

Evaluation von Managementdiagnostik: Methoden, Konstrukte, Beispiele

Sabine Hochholdinger

Evaluation von Managementdiagnostik: Methoden, Konstrukte, Beispiele

Inaugural-Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Philosophie der Universität Mannheim

vorgelegt von
Sabine Hochholdinger
aus Künzelsau (Hohenlohe)

Fakultät für Sozialwissenschaften
Dekanin: Prof. Dr. Dagmar Stahlberg
Prodekan: Prof. Dr. Jan van Deth
Referent: Prof. Dr. Werner W. Wittmann
Koreferent: Prof. Dr. Walter Bungard

Letzte mündliche Prüfung am 27.11.2002

MEINEN ELTERN
IN LIEBE

Inhaltsverzeichnis

1	Einführung	1
1.1	Hintergrund	1
1.1.1	Evaluationsforschung	2
1.1.2	Anwendungssituation	2
1.1.3	Eignungskonstrukte	3
1.2	Umsetzung	5
1.2.1	Ziele	5
1.2.2	Studien	6
2	Evaluation der Managementdiagnostik	9
2.1	Übersicht	9
2.2	Evaluation	10
2.2.1	Begriffsbestimmung	10
2.2.2	Rahmenmodelle	12
2.2.3	Umsetzung Managementdiagnostik	16
2.3	Implementierungs-Umgebung	18
2.3.1	Betrieblicher Rahmen	18
2.3.2	Einsatzziele	19
2.4	Diagnostische „Maßnahmen“	20
2.4.1	Diagnostische Urteilsbildung	21
2.4.2	Interview	22
2.4.3	Führungsfeedback	24
2.4.4	Assessment-Center	25
2.4.5	Psychometrischer Test	28
2.4.6	Computergestütztes Szenario	29
2.4.7	Arbeitsprobe	32
2.4.8	Biografische Daten	34
2.5	Fazit	36

3	Führungsforschung	37
3.1	Übersicht	37
3.2	Forschungsstrategien	38
3.2.1	Definition	38
3.2.2	Funktionale Analysen	40
3.2.3	Anforderungsanalysen	43
3.2.4	Aktivitätsstudien	45
3.2.5	Kategoriensysteme und Taxonomien	46
3.2.6	Integrative Ansätze	47
3.3	Klassische Ansätze	49
3.3.1	Traittheorien	50
3.3.2	Verhaltensorientierte Ansätze	51
3.3.3	Kognitive Ansätze	52
3.3.4	Führungsstilforschung	53
3.3.5	Interaktionistische Ansätze	54
3.3.6	Rollenansätze	56
3.4	Neuere Ansätze	57
3.4.1	Systemische Führungstheorie	57
3.4.2	Mikropolitik	60
3.4.3	Symbolische Führung	61
3.4.4	Charismatische und transformationale Führung	62
3.4.5	Subjektive Führungstheorien	65
3.5	Fazit	65
4	Führungsrelevante Eigenschaften?	67
4.1	Übersicht	67
4.2	Leistungsbezogene Eigenschaften	67
4.2.1	Intelligenz	68
4.2.2	Wissen	71
4.2.3	Komplexes Problemlösen	72
4.2.4	Weitere Konstrukte	76
4.3	Nicht leistungsbezogene Eigenschaften	79
4.3.1	Persönlichkeit	79
4.3.2	Motivation	81
4.3.3	Interpersonelle Fertigkeiten	83
4.4	Anforderungsmodell	86
4.4.1	Hauptkategorien Problemlösen	88
4.4.2	Hauptkategorie interpersoneller Bereich	89
4.5	Fazit	90

5	Outcome-Evaluation: Praxisdaten	91
5.1	Übersicht	91
5.2	Auswertung von Stellenanzeigen	92
5.2.1	Methode	92
5.2.2	Ergebnisse	94
5.3	Outcome-Evaluation im Konzern	97
5.3.1	Hintergrund	98
5.3.2	Methode	99
5.3.3	Durchführung	100
5.3.4	Ergebnisse	105
5.4	Metaanalyse: Persönlichkeit	107
5.4.1	Vorbereitung	107
5.4.2	Auswertung	110
5.4.3	Ergebnisse	111
5.5	Fazit	113
6	Treatment-Evaluation	115
6.1	Übersicht	115
6.2	Hintergrund	116
6.2.1	Praxisziele	116
6.2.2	Betriebliches Anforderungsmodell	117
6.3	Durchführung	118
6.3.1	Verfahren	118
6.3.2	Daten	121
6.3.3	Evaluationsstrategie	122
6.4	Ergebnisse	124
6.4.1	Objektivität der Urteiler	124
6.4.2	Konstruktvalidität	125
6.4.3	Kriteriumsvalidität	131
6.4.4	Gewichtung der Instrumente	134
6.4.5	Entscheidungsmodell	137
6.4.6	Feedbackprozess	138
6.5	Fazit	140
7	Impact-Evaluation	141
7.1	Übersicht	141
7.2	Hintergrund	142
7.2.1	Praxisziele und Teilnehmer	142
7.2.2	Betriebliches Anforderungsmodell	143
7.3	Durchführung	145
7.3.1	Computergestütztes Szenario	146

7.3.2	Persönlichkeitstests	148
7.3.3	Selbst- und Fremdbeurteilung	150
7.3.4	Arbeitsprobe	150
7.4	Ergebnisse	151
7.4.1	Evaluationsstrategie	152
7.4.2	Aggregation	154
7.4.3	Kriteriumsvalidität Problemlösen	157
7.4.4	Konstruktvalidität	159
7.5	Fazit	161
8	Outcome-Evaluation: Forschungsmodell	163
8.1	Übersicht	163
8.2	Methode	164
8.2.1	Anforderungsanalyse	164
8.2.2	Qualitative Inhaltsanalyse	167
8.2.3	Kodierung	170
8.2.4	Quantifizierung	174
8.3	Material	175
8.3.1	Vorbereitung und Recherche	176
8.3.2	Ausgeschlossene Literatur	178
8.3.3	Funktionale Analysen	180
8.3.4	Empirische Untersuchungen	182
8.3.5	Instrumente	184
8.4	Ergebnisse	185
8.4.1	Probelauf	186
8.4.2	Hauptkategorien	191
8.4.3	Teilkategorien	193
8.5	Fazit	198
9	Diskussion	199
	Literaturverzeichnis	209

Tabellenverzeichnis

4.1	Intelligenz und Berufserfolg	69
4.2	Problemlösen und Führungserfolg	75
4.3	Persönlichkeit und Führungserfolg	80
4.4	Anforderungsmodell	87
5.1	Auswertungskategorien Stellenanzeigen	93
5.2	Outcome-Evaluation im Konzern	99
5.3	Anforderungsmodell in der Praxis	105
5.4	Stichprobe Metaanalyse	108
5.5	Ergebnisse Metaanalyse	110
6.1	Ergebnisse Urteilterendenzen	125
6.2	Faktorstruktur der Ratingwerte im ersten Jahr	126
6.3	Faktorstruktur der Ratingwerte im zweiten Jahr	127
6.4	Faktorstruktur der MBTI-Subskalen	131
6.5	Modellierung der Auswahl im ersten Jahr	135
7.1	Übersicht über die FSYS-Skalen	147
7.2	Faktorstruktur der Arbeitsprobe	154
7.3	FSYS und Arbeitsprobe	158
7.4	FSYS, Persönlichkeit und Selbstreflexion	159
7.5	Selbstreflexion, Persönlichkeit und Arbeitsprobe	160
8.1	Kategoriensystem zur Literaturanalyse	173
8.2	Übersicht funktionaler Analysen	181
8.3	Übersicht empirischer Studien	183
8.4	Übersicht der Erhebungsinstrumente	185
8.5	Kodierung einer funktionalen Arbeit	187
8.6	Kodierung einer empirischen Studie	188
8.7	Kodierung eines Erhebungsinstruments	189

Abbildungsverzeichnis

1.1	Symmetrieprinzip bei Validierungsstudien	4
2.1	Modell der fünf Datenboxen	13
2.2	Rahmenmodell der Evaluation	14
3.1	Klassifikation von Aktivitätsstudien	42
3.2	Leadership Behavior Dimensions	53
4.1	Anforderungsmodell an Problemlöser	74
5.1	Erwartete fachliche Qualifikationen	95
5.2	Erwartete außerfachliche Kompetenzen	96
5.3	Auswahlverfahren in Deutschland	97
5.4	Perspektiven im Kompetenzmodell	106
5.5	Streuung der Stichprobenwerte	109
5.6	Konfidenzintervall der Korrelationen	112
6.1	Ablauf der internen Auswahl	120
6.2	Urteilsmodell im ersten Jahr	128
6.3	Urteilsmodell im zweiten Jahr	129
6.4	Scree-Plot der MBTI-Subskalen	132
6.5	Kriteriumsvalidierung am Assessment-Center	133
6.6	Kriteriumsvalidierung am Vorgesetztenurteil	134
6.7	Modellierung der Auswahl im zweiten Jahr	136
7.1	Verfahren und Konstrukte	152
7.2	Operationalisierte Problemlösemaße	155
7.3	Aggregate aus Selbstreflexion und Persönlichkeit	156
8.1	Ablaufschritte der Literaturanalyse	166
8.2	Eigenes Ablaufmodell der Inhaltsanalyse	169
8.3	Kodierhäufigkeiten Hauptkategorien	191
8.4	Kodierhäufigkeiten Informationsmanagement	193

8.5	Kodierhäufigkeiten Zielbildung	194
8.6	Kodierhäufigkeiten Maßnahmenplanung	195
8.7	Kodierhäufigkeiten Selbstmanagement	196
8.8	Kodierhäufigkeiten interpersoneller Bereich	197

Kapitel 1

Einführung

„The closest analog to evaluators in the private sector are management consultants.“ (Cook & Wittmann, 1998, Seite 113)

1.1 Hintergrund

In der Forschungsliteratur zur Führungskräfteeignungsdiagnostik oder Managementdiagnostik wird häufig ein Mangel an konzeptuellen und methodischen Grundlagen beklagt (Kannheiser, 1995). Genauer gesagt, sind diese Grundlagen in der Führungsforschung vorhanden, vielmehr fehlte bislang eine tragfähige Verbindung zur Eignungsdiagnostik. Deshalb erfolgt in dieser Arbeit zunächst eine Bestandsaufnahme der Managementdiagnostik, um zu erfahren, welche diagnostischen Instrumente es gibt, wie brauchbar diese sind und welche Eignungskonstrukte sich besonders für die Managementdiagnostik anbieten.

Es ist zu vermuten, dass sich auf dieser Basis eine methodische Bewertung vornehmen lässt, aber keine besonderen, managementtypischen Verfahren zu identifizieren sind, da ihre Angemessenheit auch von der Anwendungssituation und ihrem theoretischem Bezug abhängt. Diese Anwendungssituation und entsprechende Perspektiven berücksichtigen Evaluationstheorien stärker als viele methodische Beiträge zur Eignungsdiagnostik. Aus diesen Gründen bildet ein Rahmenkonzept zur Evaluation der Managementdiagnostik den ersten, methodischen Schwerpunkt dieser Arbeit, während Eignungskonstrukte und Strategien zu ihrer Entwicklung und Bewertung den zweiten, inhaltlichen Schwerpunkt darstellen. Dabei gilt eigenschaftstheoretischen Modellen und Verfahren zur Problemlösedagnostik besonderes Augenmerk.

1.1.1 Evaluationsforschung

Wie in anderen Bereichen der Anwendungsforschung ist bei der Managementdiagnostik Qualitätssicherung wichtig (Wottawa & Thierau, 1998). Nicht nur sind hier die direkten Kosten einer Fehlentscheidung, die etwa durch Abfindung und Neueinstellung entstehen, am höchsten, sondern eine Führungskraft wirkt je nach persönlicher Eignung mehr oder weniger positiv an der Entwicklung des Unternehmens mit. Deshalb ist eine systematische Bewertung verschiedener personaldiagnostischer Verfahren im Hinblick auf die Zielerreichung, also eine Evaluation, hier wie für alle anderen wirtschaftlichen Maßnahmen wichtig.

Evaluationsstudien der Diagnostik könnten neben der bekannten Kriteriumsvalidierung an beruflichen Erfolgsindikatoren (Höft, 2001) weitere Strategien beinhalten, um diagnostische Gütekriterien zu gewinnen und den Urteilsprozess zu evaluieren. Zudem sind im Anwendungskontext zusätzliche Aspekte wichtig, insbesondere sollten die Anwenderperspektive und wissenschaftliche Theorien gleichermaßen berücksichtigt werden, damit sich die Resultate besser kommunizieren lassen und eher akzeptiert werden (vergleiche Schuler, Kaufmann & Frier, 1993).

Deshalb müssen für eine solche Evaluationsforschung umfassendere Modelle gefunden werden, die gleichzeitig spezifischer auf diese Anwendung zugeschnitten sind. Ausgehend von einer kurzen Übersicht des Evaluationsbegriffs und verschiedenen Evaluationsmodellen werden im zweiten Kapitel dieser Arbeit methodische Strategien im diagnostischen Prozess in eine allgemeine Evaluationstypologie eingeordnet. Die Studien in den Kapiteln fünf bis acht sind Anwendungsbeispiele für die Managementdiagnostik aus verschiedenen Perspektiven.

1.1.2 Anwendungssituation

Direkt mit der Evaluationsforschung ist die Anwendungspraxis in einem Feld verbunden. Alle Praktiker sollten sich darin einig sein, dass eine Führungsposition ebenso wie jede andere Stelle mit dem geeignetesten Kandidaten zu besetzen sei. Davon abgesehen, unterscheiden sich bei ihnen die meisten persönlichen Voraussetzungen. Nur selten hat der Zuständige ein Psychologiestudium mit diagnostischer Vertiefung absolviert, oft liegt auch eine spezifische Ausbildung vor, die mit der Personalauswahl und -entwicklung weniger zu tun hat. Selbst wenn die Personaleignungsdiagnostik mit genuin psychologischen Verfahren wie Persönlichkeits-, Leistungs- oder Intelligenztests betrieben wird, beträgt der Anteil von Psychologen nicht mehr als 10,2 Prozent (Schulz, Schuler & Stehle, 1985).

Angesichts der von Schuler und Funke (1989) geschätzten Anzahl von 16 Millionen personeller Einzelentscheidungen, die jährlich in deutschen Unternehmen getroffen werden, zeigt diese Statistik, dass viele Unternehmen psychologische Unterstützung nicht optimal nutzen. Das bringt zum Ausdruck, wie viel Know-how das betreffende Unternehmen investieren will, beziehungsweise wie weit dort die wirtschaftliche Bedeutung einer sorgfältigen Personaldiagnostik bewusst ist.

Offensichtlich wird selten berücksichtigt, dass sich gerade die Verfahren, die mit hohem Aufwand entwickelt wurden, und die vorwiegend aus dem wissenschaftlichen Kontext stammen, in der Personaldiagnostik als besonders erfolgreich und damit ökonomisch bedeutsam erwiesen haben (Cook, 1998). Diese Fehleinschätzung liegt möglicherweise daran, dass die besonders validen Methoden eine geringere Augenscheinvalidität aufweisen. Das gilt nicht nur für die Bewerber- (Schuler, 1990), sondern ebenso für die Anwenderperspektive (Schuler, Frier & Kaufmann, 1993).

Cook (1998) weist darauf hin, dass leistungsfähige personaldiagnostische Verfahren einen entscheidenden und monetär bezifferbaren Konkurrenzvorteil eines Unternehmens darstellen. Insofern ist ein Ausbau des diagnostischen Personalsektors keine Maßnahme, die vage im Humanbereich anzusiedeln und damit eine zu vernachlässigende Größe ist, sondern sie stellt einen erheblichen und damit gewinnbringenden Technologievorteil dar. Aufgrund der hohen Relevanz sollten personaldiagnostische Verfahren auch evaluiert werden, um aussagekräftige Ergebnisse sicherzustellen.

1.1.3 Eignungskonstrukte

In der Managementdiagnostik, in der Problemlöseforschung und in der Führungsforschung wird häufig behauptet, dass Management- oder Führungsaufgaben die gleichen Anforderungen stellen wie komplexe Probleme, doch es wurde bisher kaum systematisch untersucht (Kersting, 1999). Manchmal steht im Raum, dass diese Frage durch eine Anforderungsanalyse zu klären sei, oder dass man prinzipiell untersuchen müsse, welche Anforderungen in Form eignungsdiagnostischer Konstrukte eine Managementtätigkeit stellt. Generell bleibt festzustellen, dass ein Modell fehlt, das Eignungsdiagnostik und Führungsforschung integriert (Kannheiser, 1995).

Im Zusammenhang mit der Managementdiagnostik wird in der vorliegenden Arbeit diskutiert, welche Eigenschaftskonstrukte sich prinzipiell mit Führungsleistung verbinden lassen, wofür sich das komplexe Problemlösen anbietet. Auf konzeptueller Seite zeigt die Arbeit einige Ansätze der Führungsforschung auf, die sich möglicherweise für die Eignungsdiagnostik nutzbar machen lassen. Verschiedene Strategien, Führungsansätze zu integrieren

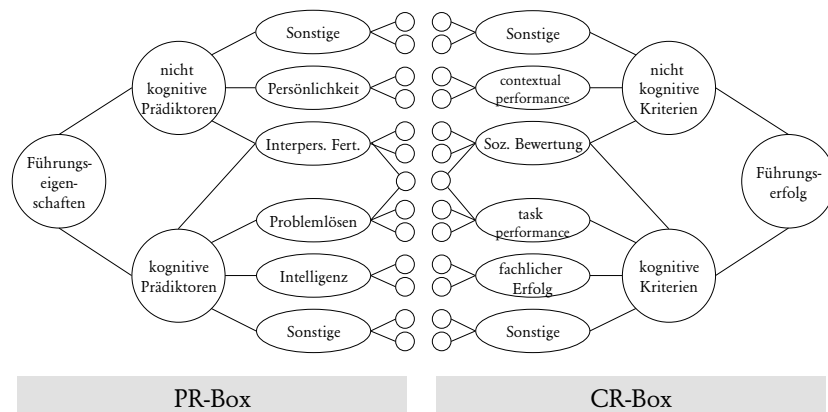


Abbildung 1.1: Symmetrieprinzip bei Validierungsstudien übertragen auf die Evaluation der Managementdiagnostik.

und mit der Eignungsdiagnostik zu verbinden, werden diskutiert. Dabei wird eine literaturgestützte Anforderungsanalyse als neue Strategie entwickelt und vor dem Hintergrund eines Problemlösemodells beispielhaft durchgeführt.

Damit verbunden erfolgt eine empirische Bewertung eines diagnostischen Instrumentariums. Besonderes Augenmerk liegt dabei auf den computer-gestützten Szenarios, mit denen gerade in der Managementdiagnostik hohe Erwartungen verbunden werden (Hasselmann, 1993). Sie werden vor allem im mittleren Management eingesetzt, und beanspruchen, mit ihrem Simulationscharakter im Sinne einer Arbeitsprobe, die Fähigkeit zum Lösen komplexer Probleme zu erfassen. Eine Validierungsstudie auf der Grundlage eines solchen Szenarios würde Hinweise darauf liefern, wieweit ein solches Szenario beruflichen Erfolg vorhersagt.

Wittmann (1990) weist darauf hin, dass bei Evaluations- und sonstigen Untersuchungen Kriterium und Prädiktor hinsichtlich Breite und inhaltlichem Bezug übereinstimmen sollen, da sonst die theoretisch maximale Korrelation zwischen Prädiktor und Kriterium von vornherein gering ist, also auch empirisch keine hohen oder mittleren Validitätskoeffizienten zu erwarten sind. Da gerade in der Anwendungsforschung und Evaluationsforschung häufig ad hoc verfügbare Kriterien mit den theoretisch interessierenden Prädiktoren korreliert werden, sollten diese Symmetrieaspekte (Brunswik, 1955) bei solchen Studien bedacht werden. Abbildung 1.1 veranschaulicht ausgewählte denkbare Kriterien und Prädiktoren auf verschiedenen Ebenen.

So ließen sich beispielsweise kognitive und nicht kognitive Prädiktoren unterscheiden. Zu den kognitiven Prädiktoren können Intelligenz, Wissen oder Problemlösefähigkeit gehören, weitere sind denkbar, etwa Arbeitsgedäch-

niskapazität. Nicht kognitive Prädiktoren könnten sich aus Persönlichkeit, interpersonellen Fertigkeiten, Motivation oder anderen Konstrukten zusammensetzen. Ihnen sollten inhaltlich adäquate Kriterien gegenüberstehen, die außerdem auf der gleichen Aggregationsebene erfasst sind. Ob ein globales Vorgesetztenurteil zum Beispiel eher kognitive oder nicht kognitive Aspekte abbildet, ist nicht zu entscheiden, wahrscheinlich beides. Damit ist für dieses Kriterium von vorneherein eine eher geringe Korrelation mit spezifischen Prädiktoren wie etwa Persönlichkeitsskalen zu erwarten.

1.2 Umsetzung

Mit den beschriebenen Problemen sind einige Punkte angesprochen, die bei der Evaluation der Managementdiagnostik eine Rolle spielen können. Es geht in dieser Arbeit vor allem darum, ein allgemeines Rahmenkonzept zunächst für die Evaluation der Managementdiagnostik vorzustellen, und damit verbundene konzeptuelle und methodische Aspekte anzusprechen, weniger um eine vollständige Modellrealisierung und Hypothesenprüfung. Die empirischen Studien stellen beispielhafte Umsetzungen für verschiedene Teilfragen dar. In der Eignungsdiagnostik überwiegen korrelative Designs, häufig mit explorativem Charakter, die mit den bekannten Einschränkungen quasi-experimenteller Designs verbunden sind.

1.2.1 Ziele

Damit verfolgt diese Arbeit mehrere Ziele, die durch verschiedene Kapitel und Studien abgedeckt werden. Generell lassen sich hier konzeptuelle Ziele, methodische Strategien und empirische Fragen unterscheiden, zu denen verschiedene Teile der Arbeit unterschiedliche Beiträge liefern.

- Aus der Programmevaluation wurde ein Modell auf die Evaluation der Managementdiagnostik übertragen, das umfassender ist als allein eine Kriteriumsvalidierung.
- Es wird ein kurzer Abriss der „Treatments“ der Managementdiagnostik gegeben, also der eignungsdiagnostischen Verfahren, damit verbunden erfolgt eine Diskussion ihrer Brauchbarkeit angesichts der Befundlage.
- Da Führungsforschung eine wichtige konzeptuelle Basis der Managementdiagnostik darstellt, ist ihr ein eigenes Kapitel mit ausgewählten Führungstheorien und Ansätzen gewidmet, um festzustellen, welche Beiträge sie zu Anforderungsmodellen in der Diagnostik liefern.

- Implizite Eignungstheorien von Praktikern als Evaluationskomponente bildet eine Auswertung von Stellenanzeigen ab.
- Eine Studie zur Zielevaluation im Konzern dokumentiert die Entwicklung eines praxisbezogenen Anforderungsmodells.
- Welche Persönlichkeitsdimensionen bei Führungskräften besonders relevant für Berufserfolg sind, wurde in einer Metaanalyse untersucht.
- Es wird gezeigt, welche Evaluationskriterien sich anhand vorliegender Daten einer Treatment-Evaluation ableiten und prüfen lassen, und welche Einschränkungen in der Praxis bestehen.
- In einer Impact-Evaluation einer Führungskräfteauswahl wird die Kriteriumsvalidität verschiedener Verfahren, insbesondere zur Erfassung von Problemlösefähigkeit und interpersonellen Fähigkeiten unter besonderer Berücksichtigung der Symmetrieprinzipien untersucht.
- Schließlich wird eine qualitative Methode zur literaturgestützten Anforderungsanalyse mit Schwerpunkt auf Problemlösen entwickelt und umgesetzt.

1.2.2 Studien

Insgesamt werden sechs Untersuchungen vorgestellt, die schwerpunktmäßig einem von drei Evaluationstypen nach Chen (1990) zuzuordnen sind, dabei aber verschiedene Fragestellungen beinhalten. Sie unterscheiden sich im empirischen Gehalt und Geltungsbereich und werden im Folgenden kurz charakterisiert. Drei empirische Studien behandeln verschiedene Aspekte der Ziel-Evaluation. Zwei weitere, jedoch umfangreichere Studien beinhalten als Schwerpunkt die Treatment- und die Impact-Evaluation. Abschließend wird ein Forschungsmodell als Beitrag zur Ziel-Evaluation präsentiert.

1. *Ziel-Evaluation der Praxisperspektive* Um eine Übersicht darüber zu gewinnen, welche fachlichen und außerfachlichen Anforderungen an Führungskräfte kommuniziert werden und welche diagnostischen Verfahren derzeit in deutschen Unternehmen bei der Auswahl der Nachwuchsführungskräfte verwendet werden, wurden 83 Stellenanzeigen ausgewertet. Im *Stellenreport 2000/2001* präsentierten sich in Deutschland ansässige Unternehmen, die hauptsächlich Führungsnachwuchs suchen, mit Stellenangeboten für Hochschulabsolventen und informieren über Anforderungen und den Ablauf der Personalauswahl.

2. *Ziel-Evaluation im Konzern* Ziel des Konzerns war es, ein Kompetenzmodell für Führungskräfte zu entwickeln und dabei Wissenschaftlichkeit mit unternehmensspezifischen Anforderungen zu verbinden. In einem mehrstufigen Prozess wurden neben dem diagnostischen Modell mehrere Perspektiven integriert. Ausgehend von dem Problemlösemodell wurde dieses erweitert um die besondere Unternehmenssituation, um die Erwartungen erfolgreicher Stelleninhaber sowie verschiedener Personalleiter im Konzern. Um den Prozess zu strukturieren, wurde zur Ziel-Evaluation die Methode der nominalen Gruppen eingesetzt.
3. *Ziel-Evaluation von Forschungsdaten* Um die Befundlage zu Persönlichkeitseigenschaften und Führungserfolg im Sinne relevanter Anforderungen systematisch zusammenzufassen, sollten hierzu neuere empirische Studien mit Führungskräften in einer Metaanalyse integriert werden, um auch einen Rückschluss auf Kompetenzmodelle ziehen zu können. Es wurden 16 Studien mit insgesamt 18 Untersuchungsstichproben ausgewertet, welche die Big-Five-Dimensionen (McCrae & Costa, 1987) sowie Führungserfolg erfassten. Für jede Persönlichkeits-Dimension lagen maximal 16 (Extraversion) und mindestens 8 (Offenheit) Korrelationskennwerte vor. Die Gesamtstichprobe umfasste 10398 Teilnehmer.
4. *Treatment-Evaluation im Konzern* Eine interne Potenzialdiagnostik mit 241 oberen Führungskräften im Konzern wurde evaluiert. Die Diagnostik erfolgte über zwei Jahre, in einem jeweils zweifach gestuften Auswahlprozess mit insgesamt fünf verschiedenen diagnostischen Verfahren. An der ersten Erhebung nahmen 157 Führungskräfte in der ersten Stufe teil, davon 81 in der zweiten Auswahlstufe. Im folgenden Jahr kamen von insgesamt 84 Führungskräften 31 in die zweite Stufe. Aufgabe der wissenschaftlichen Begleitung war es, aus den eignungsdiagnostischen Daten Evaluationsstrategien für das gesamte Verfahren zu entwickeln und mit entsprechenden statistischen Verfahren auszuwerten.
5. *Impact-Evaluation im Unternehmen* Bei einer Auswahlmaßnahme von 68 Nachwuchsführungskräften bildete ein Modell aus Persönlichkeits- und Problemlösekonstrukten die Grundlage, um Prädiktoren und Kriterien zu modellieren und erwartete Zusammenhänge korrelativ zu prüfen. Operationalisiert wurde dies durch ein computergestütztes Verfahren und verschiedene Fragebogen auf der Prädiktorseite, Kriterien waren eine fachbezogene Arbeitsprobe und Vorgesetztenurteile.
6. *Forschungsmodell Ziel-Evaluation* Eine literaturgestützten Anforderungsanalyse sollte einen Beitrag zur Evaluation der Managementdiagnostik leisten, indem sie ein Forschungsmodell für eine normative

Outcome-Evaluation erarbeitet. Deshalb wurden Beiträge aus der Führungsliteratur aus der Perspektive eines diagnostisch nutzbaren Modells zum komplexen Problemlösen und interpersonellen Fertigkeiten systematisiert und integriert. Die Literaturstichprobe enthielt 27 Arbeiten: funktionale Analysen, Aktivitätsstudien und Instrumente zur Erfassung des Führungsverhaltens, die Gesamtstichprobe umfasste 2824 Teilnehmer.

Kapitel 2

Evaluation der Managementdiagnostik

2.1 Übersicht

Die sogenannte Managementdiagnostik, also die Eignungsdiagnostik bei Führungskräften (Sarges, 1995) ist für Unternehmen höchst relevant. Jedoch wird im betrieblichen Alltag nicht alles umgesetzt, was möglich und sinnvoll ist. Das kann verschiedene Gründe haben. So sind viele diagnostische Strategien und Methoden anfällig für Urteilsfehler, wenn eine Person mehrere Informationen integriert und dabei ein klinisches Urteil (Meehl, 1954) bildet. Außerdem ist die Validität dieser Informationen nicht immer geklärt (Cook, 1998). Eignungsdiagnostische Maßnahmen in der Praxis können gelegentlich von einer idealen Diagnostik abweichen, und es ist wichtig, das zu erforschen. Gleichzeitig sollte geklärt sein, was aus wissenschaftlicher Sicht sinnvoll ist. Beide Perspektiven, die wissenschaftliche und die Anwenderperspektive, sind bei einer Evaluation managementdiagnostischer Maßnahmen zu berücksichtigen.

Unter Evaluation wurde dabei hauptsächlich Kriteriumsvalidierung verstanden. Evaluationsstudien der Diagnostik können zusätzliche Strategien zur Gewinnung diagnostischer Gütekriterien und zur Evaluierung des Urteilsprozesses beinhalten. Aus diesen Gründen müssen für eine solche Evaluationsforschung umfassendere Modelle gefunden werden. Gleichzeitig sollten sie ähnlich wie in der Trainingsevaluation (Kirkpatrick, 1987) spezifischer auf diese Anwendung zugeschnitten sein. Deshalb wurde der allgemeine Ansatz der „theory driven evaluations“ (Chen, 1990) angewendet, um verschiedene Aspekte der Managementdiagnostik zu integrieren und zudem die Perspektive der Auftraggeber zu berücksichtigen. In diesem Kapitel werden methodi-

sche Strategien im gesamten diagnostischen Prozess in die Evaluationstypologie eingeordnet. Zum betrieblichen Umfeld und Brauchbarkeit diagnostischer Verfahren für die Managementdiagnostik wird eine erste Bestandsaufnahme vorgenommen.

2.2 Evaluation

Dieser Abschnitt enthält zunächst einen knappen Abriss über wesentliche Fragen und Kontroversen der Evaluationsforschung, soweit diese relevant sind für die Managementdiagnostik, ohne Anspruch auf Vollständigkeit. Umfassendere Informationen enthalten im deutschsprachigen Raum zum Beispiel Wottawa und Thierau (1998), detaillierter Wittmann (1985) oder Rossi (1999). Als gewähltes Rahmenmodell für die Managementdiagnostik wird der Ansatz von Chen (1990) vorgestellt.

2.2.1 Begriffsbestimmung

Evaluation bedeutet allgemein die Erfolgskontrolle einer Intervention. Dabei werden das Konzept der Maßnahme, das Design, die Umsetzung und die Wirksamkeit betrachtet. Wie jede Maßnahme, die ein bestimmtes Ziel verfolgt, kann auch ein eignungsdiagnostisches Instrumentarium evaluiert werden. Wie weit oder eng der Begriff Evaluation gefasst ist, darin unterscheiden sich die meisten Publikationen (Brandtstädter, 1990). Programmevaluation im ursprünglichen Sinne ist ein Prozess, mit dem eine Maßnahme systematisch beurteilt wird. Wissenschaftliche Verfahren der Datenerhebung und -auswertung erfassen die Beurteilungskriterien, die sich in Aufwand, Effektivität und Angemessenheit gliedern. Der Evaluationsprozess kann zum Projektmanagement begleitend genutzt werden, in einen Rechenschaftsbericht münden oder für die weitere Planung verwendet werden (Wittmann, 1985).

Die klassische Programmevaluation wurde und wird im öffentlichen Raum eingesetzt, um den Erfolg sozialer Maßnahmen zu überprüfen, beispielsweise Bildungsprogramme oder andere soziale Maßnahmen, die flächendeckend Anwendung finden. In der Psychologie wird sie hauptsächlich im Zusammenhang mit Psychotherapie und Unterrichtsformen durchgeführt, aber auch in anderen Bereichen. Dennoch kann man Programmevaluation beziehungsweise Evaluationsforschung nicht als einheitliche Disziplin oder Methode ansehen. Zunächst ist ungeklärt, ob sie überhaupt eine eigene Disziplin darstellt. Nach Sechrest und Figueredo (1993) ist Evaluation keine Disziplin, sondern eine Aktivität, auch Rossi, Freeman und Lipsey (1999) betonen den interdisziplinären Charakter.

Dennoch gibt es für die Evaluationsforschung durch eigene Zeitschriften, Fachgruppen und so weiter Strukturen, wie sie für ein eigenes Fach typisch sind, und damit verbunden werden gemeinsame Eigenschaften formuliert (Shadish, 1986). Gleichzeitig lassen sich unterschiedliche Evaluationsparadigmen identifizieren, die man als Alternative oder Ergänzung betrachten könnte, deren Vertreter teilweise sehr konträr argumentieren, weil sie unterschiedliche Ziele und Bewertungskriterien implizieren. Worin Evaluation besteht, und wie sie durchgeführt werden soll, das wird unter mehreren Aspekten diskutiert.

Eine Diskussionspunkt umfasst die Frage nach der Wahl der Erhebungsverfahren, die häufig als ausschließliche Entscheidungsfrage dargestellt wird. Entweder seien quantitative Verfahren die Methode der Wahl und gleichzeitig qualitative Verfahren indiskutabel, oder umgekehrt. Die Argumente entsprechen im wesentlichen denen aus der sozialwissenschaftlichen Grundlagenforschung, zudem wird auf die besonderen Umstände der Anwendungssituation verwiesen (Cronbach, 1982).

Vertreter qualitativer Verfahren gehen davon aus, dass sie auf höhere Akzeptanz bei den Beteiligten und Betroffenen der evaluierten Maßnahme treffen, dadurch die Durchführung und Erhebung weniger behindern und außerdem unerwartete Widrigkeiten abbilden können, die im Einzelfall eine Rolle spielen. Dagegen meinen die Vertreter quantitativer Verfahren, dass gerade unter den widrigen Umständen die Datenerhebung möglichst objektiv sein soll, um die Ergebnisse nicht durch Störfaktoren zu verzerren. Es wurde auch versucht, Kriterien für den Einsatz qualitativer und quantitativer Verfahren zu formulieren (zum Beispiel Antoni, 1993) oder sie bestimmten Evaluationsphasen oder -typen zuzuordnen, aber solche Einteilungsversuche alleine greifen manchmal zu kurz.

Unmittelbar daran schließt sich die zweite Kontroverse an, ob experimentelle oder eher quasiexperimentelle Designs präferiert werden sollen (Cook & Campbell, 1979). Experimentelle Designs stellen die optimale Voraussetzung für eindeutige kausale Schlussfolgerungen dar, auch wenn sie dies nicht garantieren können. Aus der Sicht der Kritiker hingegen erfordern sie künstliche Arrangements, die sich zudem in der Anwendungssituation selten herstellen und noch seltener replizieren lassen. Quasiexperimentelle Designs sind flexibler und optimieren die externe auf Kosten der internen Validität.

Schließlich sind sich nicht alle Evaluationsforscher einig, ob es genügen kann, Evaluation aus einer rein methodischen Perspektive durchzuführen oder ob sie theoriegeleitet stattfinden soll. Sozialwissenschaftler haben meist das nötige Methodenrepertoire, um die typischen Erhebungen und Auswertungen einer beliebigen Evaluationsstudie durchzuführen. Strittig ist jedoch, ob das genügt, um den Zielen, Inhalten und Ergebnissen einer Maßnahme

gerecht zu werden, oder ob ein tieferes Verständnis der zugrundeliegenden inhaltlichen Theorien nötig ist und in die Evaluation einbezogen werden muss (Weiss, 2001).

2.2.2 Rahmenmodelle

Als ein besonderes Anwendungsfeld der Arbeits- und Organisationspsychologie hat sich die Trainingsevaluation entwickelt, teilweise in Verbindung mit der Programmevaluation, teilweise mit eigenen Modellen und Annahmen. Hierbei geht es darum, die Wirksamkeit von fachlichen und außerfachlichen Trainings oder Fortbildungen festzustellen, wobei vor allem Kriterien und Bedingungen diskutiert werden. Am bekanntesten ist dabei die Taxonomie von Kirkpatrick (1987) mit ihren vier Lernerfolgsebenen. Damit ist die Evaluationsforschung im arbeits- und organisationspsychologischen Feld vor allem in der Personalentwicklung verankert (Höft, 2001), außerdem werden inzwischen sehr viele andere Maßnahmen wie Arbeitsformen, Verbesserung betrieblicher Gesundheit und so weiter evaluiert (Holling & Gediga, 1999).

Weniger explizit als in der Trainingsevaluation ist die Evaluation bisher in der beruflichen Eignungsdiagnostik als arbeits- und organisationspsychologisches Anwendungsfeld verankert. Zum einen gibt es dabei Validierungsstudien als Strategie der Qualitätssicherung, zum anderen erscheint es nicht so leicht, Evaluationsmodelle anzuwenden, da die erwartete Struktur von Maßnahme und Erfolgsdaten nicht so offensichtlich ist. Dennoch lassen sich eignungsdiagnostische Maßnahmen nicht nur gut in die Evaluationsterminologie übertragen, sondern darüber hinaus liefern Evaluationsmodelle eine breitere Perspektive bei der Qualitätssicherung, da sie die Interessen der Auftraggeber und sonstiger Beteiligter explizit berücksichtigen.

Um Evaluationsvorhaben nach den diskutierten Aspekten zu klassifizieren, oder diese Punkte zu integrieren, wurden immer wieder Evaluations-taxonomien oder -typologien vorgeschlagen, die bestimmte Strategien beinhalten. Zu den bekannteren gehören die Einteilung in summative und formative Evaluation (Scriven, 1991) und das CIPP-Modell von Stufflebeam (2000), das die aufeinanderfolgenden Phasen Kontextevaluation, Inputevaluation, Prozess- und Programmevaluation benennt. Eine weiterer Ansatz ist das fünf Datenboxen-Modell von Wittmann (1990), das er im Zusammenhang mit seiner multivariaten Reliabilitätstheorie als Konzept zur Systematisierung von Evaluationsstudien vorstellt (Abbildung 2.1). Es soll als Idealstandard für Forschungsprogramme besser interpretierbare Ergebnisse ermöglichen.

Jede der fünf Datenboxen ist ähnlich wie der Cattellsche Datenwürfel aus den drei Dimensionen Variable, Person und Situation zusammengesetzt und

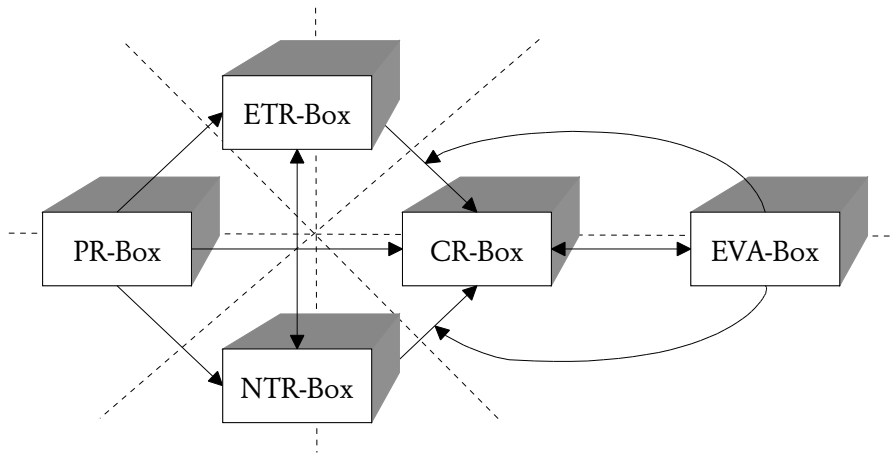


Abbildung 2.1: Modell der fünf Datenboxen (Wittmann, 1990).

steht in Beziehung zu den anderen Boxen wie in Abbildung 2.1 abgebildet. Die Prädiktorbox (PR-Box) enthält alle Daten, die bereits vor der Untersuchung vorliegen und als unabhängige Variablen die Kriterien der Evaluationsstudie vorhersagen können. Bei den Prädiktoren handelt es sich meist um demografische, organismische oder bereits angefallene Forschungsdaten. In der experimentellen Treatmentbox (ETR-Box) sind die Informationen über ein experimentelles Treatment enthalten, also eine Maßnahme, die im Rahmen eines randomisierten Kontrollgruppendesigns angewendet und evaluiert wird. So kann von den Effekten stärker auf das Treatment geschlossen werden.

Jedoch ist eine vollständige Randomisierung aus praktischen Gründen in der Anwendungsforschung nicht immer möglich, weshalb für eine nicht zufälligen Zuordnung von Treatment und Teilnehmern die nicht experimentelle Treatment-Box vorgesehen ist (NTR-Box). Die meisten Führungstrainings wären dieser Box zuzuordnen. Unter diesen Umständen kann die Zuordnung zum Treatment mit der Prädiktor-Box korreliert und damit konfundiert sein. Jedoch lässt sich dieser Varianzanteil herausrechnen, wenn die Daten der Prädiktor-Box bekannt sind, und es ist dennoch eine konservative Schätzung der Treatment-Wirkung möglich. Die Kriterienbox (CR-Box) enthält alle abhängigen Variablen beziehungsweise Indikatoren für die Wirksamkeit eines Programms im Sinne der Fragestellung. Dies können objektive und subjektive Maße sein, die zur Programmevaluation herangezogen werden.

Die Evaluationsbox (EVA-Box) schließlich umfasst alle wichtigen Forschungsfragen und Evaluationsziele, welche als Grundlage für die Untersuchung der Beziehungen zwischen den restlichen Boxen dienen. Daraus leiten

sich die Bewertungsmaßstäbe für die Evaluationsstudie ab. Meistens legen die Auftraggeber oder Stakeholder die Leitfragen für eine Evaluationsstudie fest. Aus den Beziehungen zwischen den Datenboxen lassen sich verschiedene Varianzquellen herleiten und deren Anteil prüfen, wobei diese nur sinnvoll überprüft werden können, wenn sie Symmetrieprinzipien sen. Brunswik (1955) entsprechen. Daneben werden Reliabilitätsmangel beziehungsweise -schätzungen als weitere Erklärung für geringe Effekte und Replizierbarkeit konzeptionalisiert.

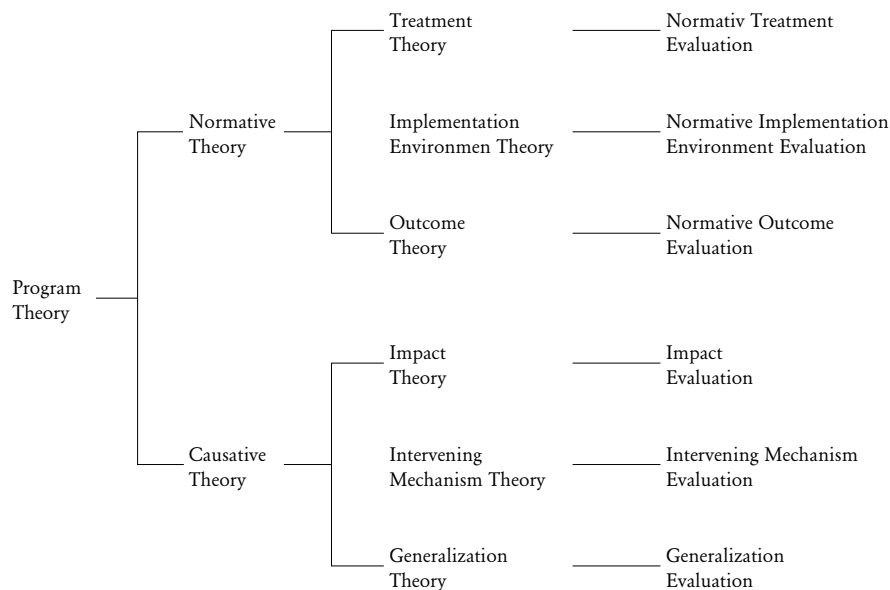


Abbildung 2.2: Rahmenmodell für die Evaluation der Managementdiagnostik (Chen, 1990).

Zeitgleich wurde die Evaluationstypologie von Chen (1990) veröffentlicht. Sie ist weniger auf spezielle psychologische und methodische Fragestellungen abgestimmt, erlaubt dagegen eine ziemlich umfassende Betrachtung des Evaluationsgeschehens, auch aus der Sicht der Managementdiagnostik. Sie integriert und erweitert die oben genannten Modelle, indem sie einzelne Studien nach Zielsetzungen unterscheidet und klassifiziert. Auch Chens Modell ermöglicht es, einzelne Studien in ein umfassenderes Forschungsprogramm zu integrieren. Jedoch klassifiziert er Studien, während Wittmanns Modell (1990) anfallende und zu berücksichtigende Informationen in Evaluationsstudien systematisiert. So sind diese beiden Modelle trotz partieller Überschneidungen keine konkurrierenden Ansätze, sondern ergänzen sich aufgrund der unterschiedlichen Ebenen, die sie ansprechen.

Insgesamt gibt es laut Chen (1990) sechs Evaluationstypen, die einzeln oder kombiniert auftreten können, und die mit ihren Zielsetzungen jeweils besondere Fragestellungen verfolgen. Die ersten drei gehören zu der Gruppe der normativen Evaluationsstudien, die unter verschiedenen Gesichtspunkten betrachten, ob die Umsetzung so stattfindet wie geplant. Die restlichen drei sind die kausalen Evaluationsstudien, bei denen aus den empirischen Ergebnisdaten Schlussfolgerungen gezogen werden sollen. Chen (1990) geht nicht davon aus, dass die beschriebenen Evaluationsfragen nur abschließend und bewertend, summativ, überprüft werden, sondern hält es für vorteilhafter, dass die Informationen wenn möglich noch formativ in der Maßnahme berücksichtigt werden, um diese möglichst optimal zu gestalten.

Das Ziel der normativen Ziel-Evaluation besteht darin, festzustellen, ob die Evaluationskriterien schlüssig, präzise und konkret sind, ob sie in einem sinnvollen Zusammenhang zu den allgemeinen Zielen stehen und ob sie operationalisierbar sind. Das kann mit den bekannten Techniken zur Entscheidungsfindung geprüft und erarbeitet werden. Mit der normativen Treatment-Evaluation wird die postulierte Struktur des Treatments mit der vorgefundenen verglichen, und eine normative Evaluation der Implementierungs-Umgebung prüft, ob die Maßnahme tatsächlich unter den intendierten Bedingungen stattfindet.

Eine Impact-Evaluation untersucht die Wirkung einer Maßnahme und ob die Wertunterschiede der Kriteriumsindikatoren eindeutig darauf zurückzuführen sind. Ob die Wirkung der Maßnahme durch zwischengeschaltete Prozesse zwischen Treatment und Erfolgsindikatoren erklärbar ist, das ist Gegenstand der Evaluation intervenierender Mechanismen. Zuletzt bezieht sich die Generalisierungs-Evaluation darauf, wie weit sich Stichproben, Maßnahmen, Indikatoren und Ergebnisse verallgemeinern lassen.

Chen (1990) hält sein Rahmenmodell der theoriegeleiteten Evaluation bewusst flexibel, was den Einsatz qualitativer und quantitativer Methoden betrifft oder die Wahl der Untersuchungsdesigns und betont, dass man ihre Gestaltung davon abhängig machen soll, wie man optimale Informationen erhält, und gegebenenfalls mehrere Zugänge geschickt kombiniert. Besonders berücksichtigt er generell den Austausch mit den Auftraggebern und anderen Beteiligten aus der Praxis, da ihre Perspektive ja eine wesentliche Rolle bei der Evaluation spielt und so die Akzeptanz der Ergebnisse verbessert wird. Schließlich ist es ein häufiges Problem, dass eine Evaluationsstudie an den Erwartungen der Auftraggeber vorbei geht.

Zu der Kontroverse um theoriegeleitete oder rein methodenorientierte Evaluation vertritt Chen jedoch eine eindeutige Position, wie sich im Titel seines Buches *theory driven evaluation* bereits ausdrückt. Demnach soll der evaluierende Wissenschaftler inhaltliche Theorien zum jeweiligen Gegen-

stand einbeziehen, da diese dem Praktiker nicht immer so bekannt sind und die Qualität der Maßnahme sowie die Aussagekraft der Evaluationsstudie steigern können. Im folgenden werden wesentliche inhaltliche Aspekte und methodische Strategien der Eignungsdiagnostik, insbesondere der Managementdiagnostik dem Rahmenmodell zugeordnet.

2.2.3 Umsetzung Managementdiagnostik

Eine **normative Treatment-Evaluation** vergleicht die postulierte Struktur des Treatments mit der vorgefundenen und bezieht sich besonders darauf, wie gut die Maßnahme im geplanten Sinne umgesetzt wurde. Die zugehörige Treatment-Theorie spezifiziert das zu evaluierende Programm, also die diagnostische Maßnahme und Beurteilung. Damit lassen sich wichtige Strategien zur Überprüfung diagnostischer Urteilsbildung und Verfahren hier einordnen. Neben allen Verfahren zur Ermittlung von Itemgütekriterien bei psychometrischen Testverfahren gehören hierzu Methoden, um die Objektivität zu ermitteln, die Reliabilität und einige Indikatoren für Validität. Objektivitäts- und Reliabilitätsindikatoren spielen insbesondere dann eine Rolle, wenn dies für den Urteilsprozess nicht als gesichert angesehen werden kann, zum Beispiel bei Ratingwerten durch Beobachter, Kollegen, Vorgesetzte oder Mitarbeiter und bei der Zusammenführung verschiedener Informationen. Hier kommen als Einschränkung der Objektivität und Fairness möglicherweise Urteilsheuristiken ins Spiel, die sich ebenfalls rechnerisch prüfen lassen.

Daneben sind im Rahmen einer Treatment-Evaluation weitere Eigenschaften des diagnostischen Prozesses zu berücksichtigen, zum Beispiel, ob die eingesetzten Verfahren nach dem Stand der Forschung für die Fragestellung und Zielgruppe überhaupt adäquat sind, oder formale äußere Kriterien, wie die Präzision und Vollständigkeit der Dokumentation und der Daten. Daneben kann, sofern möglich, qualitativ ermittelt werden, woran es liegt, wenn Objektivitäts- und Reliabilitätskennwerte ungünstig ausfallen, wenn zum Beispiel die Urteiler nicht geschult sind, was sich häufig erst auf Nachfrage herausstellt.

Eine **normative Evaluation der Implementierungs-Umgebungs** prüft, ob die Maßnahme tatsächlich unter den intendierten Bedingungen stattfindet, und die dazugehörige Theorie spezifiziert, worin diese Bedingungen bestehen, also die Implementationsumgebung. Das sind insbesondere die Teilnehmerstichprobe, die zuständigen Mitarbeiter und unternehmenspolitische und persönliche Motive aller beteiligten Personen. In der Managementdiagnostik gehört dazu die Frage ob tatsächlich die Führungsebene, für die ein Instrumentarium konzipiert wurde, auch teilnimmt, ob geschultes Perso-

nal dafür eingesetzt wird oder ob überhaupt geplant ist, die Ergebnisse der Diagnostik auch wie offiziell angekündigt zu verwenden.

Mit der **normativen Outcome- oder Ziel-Evaluation** ist festzustellen, ob die Evaluationskriterien schlüssig, präzise und konkret sind, ob sie in einem sinnvollen Zusammenhang zu den allgemeinen Zielen stehen und ob sie operationalisierbar sind. Es sollen also zunächst die Ziele der Maßnahme geklärt werden. In der Managementdiagnostik sind diese meistens in Form von Anforderungs- oder Kompetenzmodellen definiert, welche das diagnostische Instrumentarium abdecken soll. Nicht immer formulieren die Auftraggeber Erwartungen darüber, welche Leistungen es prognostizieren soll. Damit besteht ein Beitrag der wissenschaftlichen Begleitung zunächst darin, bei der Formulierung klarer, angemessener Ziele zu unterstützen und mit verschiedenen Methoden den Entscheidungsprozess zu strukturieren.

Die wissenschaftliche Begleitung kann in diesem Bereich der Evaluation als zugehörige Theorie den eigenen Forschungsstand über angemessene und bewährte Modelle zur Diagnose und Prognose von Führungserfolg einbringen. Deshalb liefern Ansätze und Untersuchungen von Modellen zur erfolgsrelevanten Beschreibung und Einschätzung von Führungsverhalten oder -Kompetenz einen Beitrag zu diesem Evaluationsbereich. Hierbei spielt außerdem eine Rolle, dass bei Aggregationsebenen und inhaltlichen Bezüge von Prädiktoren und Kriterien Symmetrieprinzipien berücksichtigt werden, um aus den Ergebnissen der kausativen Evaluationsformen auch die intendierten Schlussfolgerungen ziehen zu können (Wittmann & Matt, 1986).

Eine **kausative Impact-Evaluation** untersucht, ob die Unterschiede in Kriteriumsindikatoren eindeutig auf Unterschiede der Maßnahme zurückzuführen sind. Die Impact-Theorie enthält Hypothesen darüber, welche Eigenschaften der Maßnahme mit welchen Erfolgskriterien zusammenhängen. In der Managementdiagnostik entspricht dies den Studien zur Kriteriumsvalidierung an Berufserfolgsmaßen, wobei das Kriterium meist zeitgleich, seltener zeitlich später erfasst wird. Für die Eignungsdiagnostik besteht eine Fülle von Studien für die verschiedensten Verfahren der Eignungsdiagnostik. Neben Korrelationskoeffizienten und verwandten Maßen wird inzwischen auch der wirtschaftliche Nutzen gegenüber den Kosten von diagnostischen Verfahren im Vergleich zu Alternativen berechnet, ein wesentliches Ziel einer jeden betrieblichen Maßnahme.

Ob die Wirkung der Maßnahme durch zwischengeschaltete Prozesse zwischen Treatment und Erfolgsindikatoren erklärbar ist, bearbeitet die **kausative Evaluation intervenierender Mechanismen**. Indem die zugehörigen Theorie zugrundeliegender Prozesse spezifiziert und operationalisiert wird, kann genauer Auskunft darüber gegeben werden, in welchen Punkten die Maßnahme erfolgreich war. So können zum Beispiel mehrere Berufser-

folgskriterien erfasst werden, wie etwa objektive Leistungsdaten in Form von Umsatz, gleichzeitig Maße für den Karriereerfolg und die Mitarbeiterzufriedenheit, da möglicherweise diagnostische Persönlichkeits- und Leistungsmaße diese unterschiedlich gut vorhersagen. Außerdem lässt sich so prüfen, ob die Vorhersage über weitere, etwa motivationale oder situative Variablen beeinflusst wird. Um dies möglichst gezielt erfassen zu können, muss eine Evaluation auch hier auf inhaltliche Forschungsbefunde zurückgreifen können. So liefern die Ergebnisse der Führungsforschung Hinweise auf wichtige Variablen und Prozesse von Führungserfolg.

Zuletzt bezieht sich die **kausative Generalisierungs-Evaluation** darauf, wie weit sich Stichproben, Maßnahmen, Indikatoren und Ergebnisse verallgemeinern lassen. Die Generalisierungs-Theorie spezifiziert, worauf die Evaluationsergebnisse verallgemeinerbar sein sollen, das heißt in der Managementdiagnostik zum Beispiel, auf welche anderen Führungsebenen oder Branchen. Dazu lassen sich die bekannten Strategien der Kreuzvalidierung anwenden, daneben kann dies inhaltlich begründet werden.

2.3 Implementierungs-Umgebung

Bevor die Verfahren der Berufseignungsdiagnostik vorgestellt und diskutiert werden, sind zunächst die verschiedenen Anwendungskontexte zu beschreiben, innerhalb derer Managementdiagnostik betrieben wird. Auf der allgemeinsten Ebene wird sie von unternehmenspolitischen Entscheidungen und Strategien beeinflusst, hängt aber auch mit den Ausführenden und der Zielgruppe zusammen. Es werden Studien recherchiert, die berichten, bei welchen Zielgruppen welche Verfahren vorherrschen.

2.3.1 Betrieblicher Rahmen

Wie erfolgreich Unternehmensstrategien umgesetzt werden, hängt von den planenden und ausführenden Personen und damit von deren Befähigung ab (Laukamm & Walsh, 1985). Der Begriff des strategischen Personalmanagement beinhaltet und betont die Einbindung des Personalbereiches in die Unternehmensstrategie und in die abgeleiteten Pläne. Umgekehrt müssen Personalfachkräften die operativen Geschäfte des Unternehmens gut kennen sowie die strategische Planung. Dabei ist die Personaleignungsdiagnostik wiederum nur ein Teilgebiet der Personalarbeit. Sie umfasst außerdem Verwaltung, Entlohnung, Arbeitsrecht, Personalbeschaffung, Weiterbildung und andere Aufgabenfelder.

Dennoch sollen die eignungsdiagnostischen Maßnahmen im Einklang mit

den Unternehmensstrategien stehen. Um ein strategiekonformes Beurteilungswesen zu erreichen, müssen aus den Zielen der strategischen Geschäftseinheiten die Ziele für die Ressort-Einheiten abgeleitet werden (Jochmann, 1992). Ressort-Einheiten sind zum Beispiel Vertrieb, Marketing oder Personal. Das Ressort Personal formuliert die Anforderungen an Mitarbeiterqualifikationen, welche die zuständigen Mitarbeiter oder Abteilungen trainieren, diagnostizieren oder auswählen. Zur Diagnostik werden die entsprechenden Verfahren bereitgestellt. Welche das sind, hängt davon ab, wie lange sie dauern, wie teuer sie sind und ob wichtige Interessengruppen sie akzeptieren. Außerdem ist wichtig, wie flexibel sie sich gestalten und in den Ablauf integrieren lassen.

Neben den Unternehmensstrategien prägen personalpolitische Grundsätze die Eignungsdiagnostik. Generell entscheidet die Personalpolitik, wie Unternehmensressourcen und Personal einander zugeordnet werden, und welche Mittel im allgemeinen die Personalarbeit prägen. So unterscheiden sich Unternehmen darin, ob sie eher externe oder interne Wege der Personalauswahl bevorzugen, ob sie solche und andere Personaldienstleistungen über eine eigene Abteilung oder externe Beratungen abdecken, welche Ansprüche sie an Personalarbeit haben, und wie wichtig Personalförderung ist.

Weiterhin entscheidet die Personalpolitik darüber, wie oft welche Verfahren zu welchem Zweck eingesetzt werden. Außerdem resultiert daraus, wie sich die Personalarbeit aufgliedert, wie die Zuständigen zusammenarbeiten, und welche Qualifikationen sie haben. In Unternehmen führen entweder geschulte Laien mit nicht-psychologischem Schwerpunkt oder betriebspsychologische Mitarbeiter eignungsdiagnostische Untersuchungen durch (Meffert & Wagner, 1992).

Bei Führungskräften ist die Zumutbarkeit ein wichtiges Kriterium für die Entscheidung für ein Auswahlverfahren (Niederfeichtner, 1987), weshalb oberste Führungskräfte eher über informelle Strategien wie Referenzen, Empfehlungen und persönliche Gespräche rekrutiert und beurteilt werden, ähnlich auch in den USA (Marsden, 1994). Die Zumutbarkeit betrifft neben den eingesetzten Verfahren die beteiligten Personen. Oft wird es nicht als zumutbar für obere und oberste Führungskräfte angesehen, von Rangniederen, wie es die Mitarbeiter der Personalabteilung in der Regel sind, beurteilt und ausgewählt zu werden, oder den gleichen Verfahren wie Rangniedere unterzogen zu werden.

2.3.2 Einsatzziele

Eignungsdiagnostische Beurteilungen lassen sich zu verschiedenen Zwecken durchführen, wobei es immer darum geht, individuelle Arbeitsleistungen im

Sinne des Unternehmens zu optimieren. Diese Einsatzzwecke lassen sich als psychologische Interventionen nach der bekannten Klassifikation von Pawlik (1976) einteilen, die er nach Ansatzpunkt und Strategie unterscheidet. Der Ansatzpunkt kann die Person oder die äußere Bedingung sein, die Strategie entweder Auswahl (Selektion) oder Veränderung (Modifikation). Daraus resultieren vier Kombinationen.

Die Personalauswahl wäre eine Selektion einer oder weniger Personen aus mehreren verfügbaren Kandidaten für eine bestimmte Bedingung, die in einer Position, einer Laufbahn oder einer Förderungsmaßnahme bestehen kann. Die nicht ausgewählten Kandidaten werden abgelehnt. Bei der Platzierung oder multiplen Selektion werden alle Kandidaten jeweils aus einer Menge von Bedingungen der für sie am besten passenden zugeordnet, wie beispielsweise unterschiedlichen Entwicklungsplänen oder verfügbaren Positionen. Dabei wird niemand abgelehnt. Im Falle einer Verhaltensmodifikation sollen Personenmerkmale verändert werden, beispielsweise durch Personalentwicklungsmaßnahmen. Diese Merkmale müssen diagnostiziert werden, um Veränderungsbedarf und Entwicklungspotenzial festzustellen. Weniger wichtig für die Eignungsdiagnostik ist das vierte Feld dieses Schemas, die Bedingungsmodifikation. Hier geht es um die Veränderung von äußeren Arbeitsbedingungen.

Welche betriebliche Relevanz die Managementdiagnostik hat, zeigt sich daran, wie oft sie in der Praxis durchgeführt wird. Verschiedene empirische Erhebungen zeigen, dass Führungskräfte der mittleren und unteren Ebene sowie Nachwuchsführungskräfte die wichtigste Zielgruppe eignungsdiagnostischer Verfahren und sonstiger Maßnahmen zur Auswahl und Rekrutierung darstellen (Schuler, Frier & Kaufmann, 1993).

Insbesondere das Assessment-Center, das zeit- und personalaufwendigste Verfahren, wird in deutschen Unternehmen hauptsächlich für die externe und interne Auswahl von Trainees sowie für untere, mittlere und sogar obere Führungskräfte eingesetzt, selten für Beschäftigte ohne Führungsaufgaben. Weniger häufig als bei anderen Gruppen kommen bei Führungskräften Persönlichkeits- und Leistungstests zum Einsatz (Bartels, 1987). Aktuellere Daten wurden für diese Arbeit selbst ermittelt und im Kapitel 5 berichtet.

2.4 Diagnostische „Maßnahmen“

In der Managementdiagnostik stellen eignungsdiagnostische Verfahren das Treatment oder die Maßnahme dar. Deshalb sollte die Befundlage dazu berücksichtigt werden. Typischerweise kann man zunächst abklären, welche Informationen zu den klassischen Testgütekriterien vorliegen. Dabei ist

weiterhin zu berücksichtigen, auf welchen Daten diese Ergebnisse beruhen, insbesondere vor welchem methodischen und inhaltlichem Hintergrund Prädiktoren und Kriterien bei Validierungsstudien stehen, um diese Befunde sinnvoll interpretieren zu können. Neben inhaltlichen Beziehungen (Schuler, 1996) sind Aggregations- und Symmetrieebenen wichtig (Wittmann & Matt, 1986).

2.4.1 Diagnostische Urteilsbildung

Nach oder schon während der Erhebung müssen die gewonnenen Daten für die geplante diagnostische Aussage kombiniert und aggregiert werden. Das Vorgehen und die Interpretation hängt einerseits von der Zielsetzung der Maßnahme ab, die hauptsächlich Auswahl oder Entwicklung, andererseits vom Konstruktmodell und seiner Beziehung zur Anwendung. In diesem Rahmen wird neben den oben beschriebenen diagnostischen Entscheidungsmodellen nach Pawlik (1976) zwischen Status- und Prozessdiagnostik unterschieden. Statusdiagnostik richtet sich auf Eigenschaftskonstrukte, die relativ stabil und verfügbar sind, und steht damit bei der Personenselektion im Mittelpunkt, während Prozessdiagnostik sich eher auf veränderliche Personenmerkmale bezieht und eine Prognose über zukünftige Veränderungen insbesondere im Zusammenhang mit geplanten Interventionen abgeben soll.

Für jede diagnostische Entscheidung benötigt man ein Modell, aus dem sich Zuordnungsregeln von diagnostischen Informationen zu Konstrukten ableiten lassen, zum Beispiel die Anforderungs-Aufgaben-Matrix beim Assessment-Center. Weiterhin sollten Gewicht und Ausprägung der Prädiktoren zur optimalen Vorhersage der Kriterien festgelegt werden. Für den Fall der Zuweisung von Personen zu Bedingungen sind zusätzlich kritische Werte der einzelnen Prädiktoren (multiple Cutoffs) und alternativ oder kombiniert damit kritische Werte des aggregierten Gesamtwertes (einfacher Cutoff) festzulegen. Dabei haben sich lineare Prognosemodelle bisher als die robustesten erwiesen.

Allerdings liegt in der personaldiagnostischen Praxis nicht immer ein Zuordnungsmodell vor, oder die diagnostischen Entscheidungen beruhen manchmal auf intuitiv statt rechnerisch zusammengefassten Daten. Dadurch werden die diagnostischen Informationen nicht optimal ausgeschöpft, und es kommt hier wie in anderen Situationen zu einer klinischen Urteilsbildung, die von verschiedenen Urteils- und Entscheidungsfehlern betroffen ist (Jäger, 1982). Zu viele Informationen können nicht mehr vollständig integriert werden, und es kommt zu systematischen Verzerrungen oder Urteilsheuristiken. Daraus kann selbst bei sorgfältigster Diagnostik kein valides Urteil resultieren, wie zum Beispiel bei Einstellungsempfehlungen von Experten,

die klinisch gebildet wurden und mit der späteren Vorgesetztenbeurteilung nur zu $r=0.17$ korrelierten (Reilly & Chao, 1982).

Unter diesen Umständen wäre eine Verbesserung der beruflichen Eignungsdiagnostik und damit auch der Managementdiagnostik bereits allein dadurch zu erreichen, dass die diagnostische Aussage mittels statistischer Urteilsmodelle konsequent systematisiert wird, selbst wenn kein Anforderungsmodell vorliegt. Dies ist nach dem Stand der Forschung mit einfachen Mitteln möglich, stößt aber in der Praxis gelegentlich auf Widerstand, wenn dort eine große Wertschätzung der intuitiven Urteilsbildung besteht.

2.4.2 Interview

Neben der Auswertung der schriftlichen Bewerbungsunterlagen ist das Interview nach wie vor die häufigste Methode in Unternehmen. Allgemein versteht man darunter die zielgerichtete mündliche Kommunikation zwischen einem oder mehreren Interviewern und einem oder mehreren Befragten, wobei Informationen über das Verhalten und Erleben der befragten Person gesammelt werden sollen (Keffler, 1988). In der Eignungsdiagnostik soll es für den Interviewer die Passung von Bewerbern klären und außerdem den Bewerber informieren.

Daneben umfasst das Interview viele verschiedene Varianten. Es unterscheidet sich nach der Anzahl von Interviewern und Befragten, nach der Strukturiertheit des Gesprächsablaufes, nach der Funktion und Position der Interviewer sowie nach der Art der Urteilsbildung. Daneben dient es unterschiedlichen Zwecken. Zunächst kann es zur einfachen Feststellung von Fakten eingesetzt werden, welche die schriftlichen Unterlagen ergänzen.

Weiterhin dient es der Bewertung von Fähigkeiten, Fertigkeiten und Motivation des Kandidaten bezüglich einer bestimmten Position oder Laufbahn. Als Breitbanddiagnostikum soll es potenzielle Schwachstellen und Stärken als Grundlage weiterer Maßnahmen identifizieren. Deshalb beziehen sich die Fragen auf fachliche Qualifikationen, Fähigkeiten und Persönlichkeitseigenschaften sowie persönliche Hintergründe des Kandidaten. Eine letzte Facette ist die Art der Urteilsbildung. Entweder fasst der Interviewer seine Eindrücke zu einem klinischen Urteil, also in intuitiver Kombination und Gewichtung zu einem Gesamtergebnis zusammen, oder er bewertet, gewichtet und aggregiert die Informationen nach einem vorgegebenen Schema, das heißt er bildet ein statistisches Urteil (Meehl, 1954).

Obwohl Anwender und Betroffenen das Interview positiv bewerten und es eine große Rolle in der Auswahlentscheidung spielt, war die empirische Bewährung in früheren Validitätsstudien und Metaanalysen gering. Die Korrelation mit beruflichen Leistungskriterien lag nahe Null, wobei die Einzelbe-

funde untereinander stark variierten. Das regte angesichts der Heterogenität des Verfahrens dazu an, diese Unterschiede durch besondere Eigenschaften des Interviews aufzuklären und methodische Verbesserungen vorzunehmen.

Umso strukturierter und anwendungsnäher die Fragen sind, desto höher ist die Kriteriumsvalidität. In der Metaanalyse von Wiesner und Cronshaw (1988) lag der Validitätskoeffizient für frei geführte Interviews bei $r=0.13$, für strukturierte dagegen betrug er $r=0.40$. Für den Zusammenhang zwischen der Strukturiertheit und der Validität von Auswahlgesprächen wurde eine Korrelation von $r=0.45$ ermittelt. Auch ist es besser, wenn es auf einer Anforderungsanalyse basiert. Damit zählt das Interview in seiner optimalen Durchführungsvariante zu den besten eignungsdiagnostischen Verfahren.

Laut Schuler (1989) verbessert sich die Kriteriumsvalidität auch, wenn Informationssammlung und Entscheidung getrennt werden, trainierte Interviewer teilnehmen und geprüfte und verankerte Skalen zum Einsatz kommen. Um die Validität zu steigern, präsentiert das situative Interview (Latham, Saari, Pursell & Campion, 1980) im Rahmen der Critical incidents technique typische Situationen des Arbeitsablaufes, für die der Kandidat eine Handlungsstrategie auswählen oder entwickeln soll, was einer Arbeitsprobe recht nahe kommt. Im multimodalen Interview (Schuler, 1992) sind typische Interviewinhalte expliziert, getrennt und in ihrer Abfolge gegliedert, die sich aus sieben Komponenten zusammensetzt.

Damit ist das Interview ähnlich wie die Arbeitsprobe unter günstigen Umständen ein brauchbares Verfahren, das hängt von seinen formalen Eigenschaften und Inhalten ab. So befassen sich einige Studien damit, inwieweit und unter welchen Umständen Interviewscores mit kognitiven Fähigkeitstests korrelieren und außerdem inkrementelle Validität bezüglich beruflicher Bewährung aufweisen (Huffcutt, Roth & McDaniel, 1996). Interviews können auch genutzt werden, um anhand der Antworten und des nonverbalen Verhaltens auf Persönlichkeitseigenschaften zu schließen (Motowidlo, Burnett, Maczynski, Witkowski, Wojtachnio & Chelapa, 1996).

Schließlich wurde die inkrementelle Validität von Interviewratings auch metaanalytisch untersucht (Cortina, Goldstein, Payne, Davison & Gilliland, 2000), wobei diese über die Interkorrelation mit Persönlichkeits- und Intelligenzmaßen hinaus einen substanziellen Beitrag zur Vorhersage von Berufserfolg leisten können, und das umso mehr, je stärker die Interviews strukturiert sind. Offen bleibt jedoch die Frage, welches psychologische Konstrukt Interviews damit erfassen, möglicherweise berufsbezogenes Wissen oder Berufserfahrung.

2.4.3 Führungsfeedback

Beurteilungsverfahren werden in der Managementdiagnostik häufig auch als Führungsfeedback, Performance Assessment oder Leadership Assessment bezeichnet. Das sind mehr oder weniger stark formalisierte Techniken, die Leistungen oder Fähigkeiten auf der Verhaltens- oder Eigenschaftsebene beurteilen sollen. Diese lassen sich im wissenschaftlichen Kontext als Kriterium verwenden, sofern berufliche Leistungen abgebildet werden (Arvey & Murphy, 1998). Häufiger dienen sie als weiterer Indikator auf Prädiktorseite zur Diagnose, Prognose der beruflichen Kompetenz oder im betrieblichen Setting zur Leistungskontrolle (Steinhoff, 1995). Generell können entweder die Beurteilten selbst diese Einschätzungen vornehmen oder andere Personen, von denen eine gute Kenntnis der Beurteilten erwartet wird, wie Vorgesetzte, Kollegen und Mitarbeiter (Runde, Kirschbaum & Wübbelmann, 2001).

Stärker formalisierte Verfahren sind nach Schuler und Funke (1993) verschiedenen Methoden zuzuordnen. Bei der freien Eindrucksschilderung werden Eindrücke über erfolgsrelevantes Verhalten oder Eigenschaften und Kompetenzen anhand eines Leitfadens in freien Worten festgehalten. Durch die geringe Standardisiertheit lässt sich das Verfahren schnell und flexibel konstruieren, worunter als Nachteil die Vergleichbarkeit der Ergebnisse leidet. Dennoch ist es am weitesten verbreitet.

Beim Einstufungsverfahren werden Verhaltensbeobachtungen, Merkmals- oder Ergebniseinschätzungen einer mehrstufigen Skala zugeordnet. Diese kann verbal verankert sein, oder numerische Ratings auf Beurteilungsdimensionen erlauben. Die verhaltensverankerten Skalen basieren häufig auf Arbeitsanalysen und wollen nur erfolgsrelevantes Verhalten in die Beurteilung einbeziehen. Seltener kommt ein Kennzeichnungs- und Auswahlverfahren zum Einsatz. Dies soll erfolgsrelevante Verhaltensweisen konkret beschreiben und Urteilstendenzen vermeiden.

Ein Rangordnungsverfahren wird vor allem im informellen Gebrauch verwendet. Dabei werden die beruflichen Leistungen der Beurteilten global oder für mehrere Leistungsbereiche eingestuft, indem diese mit anderen Beurteilten verglichen werden. Neben einfachen Rangreihen lassen sich Paarvergleiche durchführen oder Prozenstränge vergeben. Dieses Verfahren wurde hauptsächlich unter dem Aspekt der Urteilsfehler diskutiert und untersucht. Insbesondere können Urteiler häufig die Urteilsdimensionen nicht unterscheiden, sondern neigen zu Korrelationstendenzen. Ob dies durch die Verwendung von eindeutigen, verhaltensbezogenen und operational abgrenzbaren Urteilskriterien zu vermeiden ist, bleibt strittig, ebenso wie die Wirksamkeit von Beurteilertrainings.

Weiterhin ist die geringe Korrelation von Selbst- und Fremdbeurteilungen

sowie die Interkorrelation von Fremdbeurteilungen problematisch. Dennoch ist nach der Metaanalyse von Hunter und Hunter (1984) die prognostische Validität von Kollegenurteilen mit der von Vorgesetztenurteilen vergleichbar. Zu einem ähnlichen Ergebnis kommen Schmitt, Gooding, Noe und Kirsch (1984) mit einem Validitätskoeffizienten von $r=0.43$. Andere metaanalytische Studien ziehen den Schluss, dass die Prognosegüte davon abhängt, was unter beruflicher Leistung verstanden wird.

Insbesondere, wenn Führungskräfte beurteilt werden, können sich bei unterschiedlichen Perspektiven, wie sie etwa die 360-Grad-Analyse beinhaltet, die Urteile zwischen den Ratern erheblich unterscheiden (Conway & Huffcutt, 1999). Neben den psychometrischen Problemen kritisiert Neuberger (2000), dass hier die Beurteiler keine neutralen Beobachter des Verhaltens seien, sondern vor allem ihre persönlichen Erwartungen transportieren, um das Instrument zu ihren Zwecken zu benutzen.

Ohnehin tritt das Beurteilungsverfahren selten als Einzelverfahren auf, sondern meist kombiniert mit weiteren Verfahren und Maßnahmen. Nicht immer dienen Beurteilungsverfahren allein als Prädiktor beruflicher Leistung, sondern die Resultate werden in wissenschaftlichen Untersuchungen als Kriterium der beruflichen Bewährung betrachtet. In beiden Fällen wird dies mit Modellen der beruflichen Leistung eng verbunden, die für Führungsaufgaben auch mit dem Ziel einer präziseren Diagnostik zunehmend weiterentwickelt werden (Conway, 1999; Tett, Guterman, Bleier, & Murphy, 2000).

Neben der Übertragung des Fünffaktorenmodells der Persönlichkeit auf Führung (Hogan, Curphy & Hogan, 1994) gehen zum Beispiel Borman und Motowidlo (1997) von zwei Dimensionen des Führungserfolges aus, die an frühere Ansätze erinnern. Sie unterscheiden eine aufgabenunspezifische Dimension, die sogenannte Contextual performance und eine traditionelle, aufgabenbezogene Leistungskomponente, die Task performance. Contextual performance umfasst persönliche Eigenschaften wie Begeisterungsfähigkeit, prosoziales Verhalten und Spontaneität, welche zwischenmenschliche Beziehungen erleichtern und damit den Unternehmenszielen dienen. Task performance bezieht sich auf alle mit der Tätigkeit verbundenen Fähigkeiten und Fertigkeiten und auf wesentliche Techniken wie Delegation. Diese Zweiteilung ermöglicht außerdem eine kontextübergreifende Untersuchung des Führungsverhaltens.

2.4.4 Assessment-Center

Unter den Begriff des Assessment-Centers fällt eine Kombination verschiedener eignungsdiagnostischer Verfahren. Sie sollen eine umfassende Einschätzung aktueller Kompetenzen oder die Prognose zukünftiger Entwick-

lungen ermöglichen, womit Personalauswahl und -entwicklung die häufigsten Anwendungszwecke des Assessment-Centers sind. Die wichtigste manifeste Zielsetzung des Assessment-Centers ist nach Schuler (1996) die Auswahl oder Förderung von Führungskräften.

Nach Jeserich (1981) ist ein Assessment-Center in Anlehnung an die amerikanische Begriffsbestimmung ein systematisches Verfahren, das von mehreren Beobachtern gleichzeitig für mehrere Teilnehmer in Bezug auf vorher definierte Anforderungen angewandt wird. Es dient der qualifizierten Feststellung von Verhaltensleistungen. Daran nehmen meist sechs bis zwölf Beurteilte gleichzeitig teil und etwa halb so viele Beurteiler, vor allem Linien-vorgesetzte, Psychologen und Mitarbeiter der Personalabteilung.

Mehr Systematik erhält das Verfahren durch festgelegten Bewertungsregeln, durch zeitlicher Trennung von Beobachtung, Bewertung und Beobachtungserschulung. Verschiedene eignungsdiagnostische Verfahren erfassen idealerweise zum Teil dieselben Anforderungsmerkmale (Multimodalität) und sind realitätsnah gestaltet. Diese Bausteine sollen in erster Linie den Berufsalltag simulieren, wie etwa Präsentationen, Durcharbeiten eines Postkorbs, Entscheidungsprobleme, Fallstudien, Gespräche und Arbeitsgruppen. Ergänzend können psychometrische Persönlichkeits- und Leistungstests hinzukommen, Beurteilungsverfahren und Planspiele mit oder ohne Computerunterstützung.

Da die Verwendung der einzelnen Bausteine optional ist und ein Assessment-Center für den jeweiligen Einsatzzweck konzipiert wird, gibt es kein typisches Assessment-Center, sondern die Bezeichnung ist ein Sammelbegriff für sehr heterogene Durchführungsvarianten. Diese unterscheiden sich weiterhin nach der Bewertungsmethode, nach der Gesamtdauer und nach der Zuordnung von Beobachtern und Beurteilten. Deshalb kann man es nicht global bewerten, außerdem unterscheiden sich Assessment-Center darin, wie sorgfältig sie konstruiert, durchgeführt und ausgewertet werden.

Die Reliabilität eines Assessment-Centers wird üblicherweise über die Interrater-Reliabilität geschätzt, also über die Übereinstimmung mehrerer Beobachter bezüglich der Einzel- oder Gesamturteile desselben Teilnehmers. Diese liegt meist im akzeptablen Bereich und steigt, wenn sich die Beobachter zuvor über die Teilnehmer ausgetauscht haben. Gesamturteile sind außerdem höher interkorreliert als die Bewertung einzelner Bausteine. Sehr selten erfolgt eine Berechnung der Retest-Reliabilität, da gerade im betrieblichen Kontext die Beurteilten meist nur einmal am gleichen Assessment-Center teilnehmen. In den wenigen Studien, in denen Teilnehmer zweimal das gleiche Assessment-Center durchliefen, oder eine vergleichbare Parallelversion gefahren wurde, resultierten mittlere Retest-Reliabilitätskoeffizienten.

Das hängt vor allem von der letzten Phase der diagnostischen Urteils-

bildung ab, weil dabei die Einzelratings zu einem globaleren Urteil auf den vorher festgelegten Anforderungsbereichen zusammengefasst werden. Das geschieht häufig als klinisches Urteil (Meehl, 1954) und führt zu den bekannten Problemen, wie unzureichende Nutzung der vorliegenden Daten und subjektive Gewichtung einzelner oder irrelevanter Beobachtungen. Allerdings stößt ein statistisches Modell der Urteilsbildung bei vielen Beobachtern auf mangelnde Akzeptanz, insbesondere bei den Führungskräften (Jones, Herriot, Long & Drakely, 1991).

Kleinmann (1997) diskutiert die Validität des Assessment-Centers unter dem Gesichtspunkt der prognostischen Validität, die vor allem für den Anwendungszweck der Personalauswahl zentral sei, sowie unter dem Aspekt der Konstruktvalidität, die für Personalentwicklungszwecke das entscheidende Gütemaß sei. Die Validität im Sinne einer Korrelation mit Berufserfolgskriterien wurde dabei in verschiedenen Validitätsstudien und Metaanalysen geschätzt, wobei in den Einzelstudien eine hohe Streubreite vorliegt.

Nach der Metaanalyse von Thornton, Gaugler, Rosenthal und Bentson (1992) beträgt die korrigierte mittlere Validität der 50 einbezogenen Studien $r=0.37$. Ähnliche Werte werden auch in anderen Metaanalysen berichtet. Einerseits gehört damit das Assessment-Center zu den besseren eignungsdiagnostischen Verfahren, andererseits liegt dieser Wert auch nicht höher, als ihn einige der hier berichteten Einzelverfahren wie etwa Intelligenztests oder strukturierte Interviews mit wesentlich geringerem zeitlichen, finanziellen und personellen Aufwand erreichen. Problematisch sind bei den Validitätsstudien die verwendeten Kriterien, die häufiger in der Einschätzung des Potenzials bestehen als in konkreten Leistungskriterien. Für Potenzialeinschätzungen sind die Validitätskoeffizienten höher als für Leistungsdaten.

Die Konstruktvalidität ist beispielsweise über eine Multitrait-Multimethod-Validierung im Sinne von Campbell und Fiske (1959) prüfbar, da aufgrund der Multimodalität des Assessment-Centers mehrere Anforderungsdimensionen beziehungsweise Konstrukte jeweils mit mehreren Verfahren abgedeckt werden. Dabei müssten verschiedene Methoden zur Messung desselben Konstrukts hoch miteinander korreliert sein (konvergente Validität), während die Erfassung verschiedener Konstrukte durch dieselben Methoden geringere Zusammenhänge aufweisen sollten (diskriminante Validität).

Das kann durch Korrelationen oder faktorenanalytisch bestimmt werden. Allerdings ist beides den meisten Untersuchungen zufolge beim Assessment-Center nicht gegeben (Sackett & Dreher, 1982; Russell, 1987), sondern vielmehr werden die Beobachterratings durch die Effekte der Übungen dominiert, während die Unterschiede der Konstrukte kaum zum Tragen kommen. Damit kann dem Assessment-Center zwar prädiktive, aber keine Konstruktvalidität attestiert werden.

Möglicherweise ist der Grund dafür, dass die Beobachter bereits nach denselben Kriterien urteilen, nach denen später die Potenzialeinschätzung oder Leistungsbewertung auf Kriterienseite vorgenommen wird, obwohl dies nicht vorgesehen ist. Diese und andere Einflussfaktoren auf die Konstruktvalidität überprüft Kleinmann (1997) mit eigenen empirischen Studien, welche solche Schwierigkeiten systematisch beleuchten. Eine weitere methodische Kritik bezieht sich auf die Störeffekte, die unter dem Begriff der Sozialpsychologie des Experiments bekannt sind (Bungard, 1987).

Das Dimensionalitätsproblem wurde oft repliziert und diskutiert (Fleenor, 1996; Carless & Allwood, 1997), auch andere Erklärungs- oder Lösungsansätze führten nicht weiter. So wurde beispielsweise die Art der Urteilsdimensionen variiert, indem neben den klassischen Assessment-Center-Dimensionen das Abschneiden auf Dimensionen von Management-Funktionen eingeschätzt wurde (Joyce, Thayer & Pond, 1994), oder neben den grafischen Checklisten verhaltensorientierte Ratings eingesetzt wurde (Donahue, Truxillo, Conrwell & Gerrity, 1997), wobei in beiden Fälle weder divergente noch konvergente Konstruktvalidität gezeigt oder verbessert werden konnte.

Erstaunlicherweise konnte eine neuere Studie (Arthur, Woehr & Maldegen, 2000) sowohl diskriminante als auch konvergente Validität im Rahmen einer Multitrait-Multimethod-Validierung zeigen. Die Autoren finden nach eigener Aussage keine plausible Erklärung dafür, warum dies in ihrem Fall gelungen oder warum es in den bisherigen Untersuchungen fehlgeschlagen ist, wobei Verfahren mit nachgewiesener Kriteriumsvalidität logisch zwingend ebenso Konstruktvalidität aufweisen müsse.

2.4.5 Psychometrischer Test

Ein psychometrischer oder psychologischer Test ist ein standardisiertes, routinemäßig anwendbares Verfahren zur Messung individueller Verhaltensmerkmale, aus denen bei den betreffenden Personen Schlüsse auf Eigenschaften oder auf Verhaltensweisen in anderen Situationen gezogen werden. Diese Definition beinhaltet gleichzeitig zwei Aspekte der Validität, das prognostische und das diagnostische Schließen. Generell werden Leistungs- und Persönlichkeitstests unterschieden. Schuler (1996) weist explizit darauf hin, dass psychologische Testverfahren aufgrund ihrer weit entwickelten Methodologie der Testkonstruktion in der Eignungsdiagnostik zu wissenschaftlichen Zwecken die häufigsten Messverfahren sind.

Mit dem hohen Grad an Standardisierung bei Durchführung, Auswertung und Interpretation in getrennten Schritten, der den verzerrenden Einfluss von Urteilsfehlern minimiert, genügt der psychometrische Test dem Anspruch der Objektivität und wird bei der Entwicklung hinsichtlich Reliabilität und Va-

lidität optimiert. Damit setzt der psychologische Test Standards für alle anderen eignungsdiagnostischen Verfahren, auch für die Ökonomie bei der Durchführung. Für die Eignungsdiagnostik wird jedoch nicht die Brauchbarkeit eines psychometrischen Tests im Allgemeinen diskutiert und untersucht, sondern man unterscheidet die Tests nach den erfassten Inhalten. Dies wird im Zusammenhang mit führungsrelevanten Eigenschaften diskutiert.

2.4.6 Computergestütztes Szenario

Das computergestützte komplexe Szenario ist in der betrieblichen Eignungsdiagnostik ebenso wie in der kognitions- und differenzialpsychologischen Forschung inzwischen recht verbreitet (Kluwe, 1995). Es handelt sich dabei um ein Computerprogramm, das als Modell bestimmter Realitätsbereiche dient, indem es deren Strukturen und Verhalten nachbildet. Diese Realitätsbereiche sind meistens technische, ökologische oder ökonomische Systeme. In der Problemlöseforschung kann sich ein solches Szenario von einer nicht computergestützten Problemsituation darin unterscheiden, dass es komplexer ist, weniger transparent, dynamisch und vernetzt, und dass Eingriffseffekte verzögert sind.

Darüber hinaus lässt sich wenig Allgemeingültiges über Eigenschaften des computergestützten Szenarios sagen, da der Begriff sehr verschiedene Verfahren umfasst. Sie können sich zum Beispiel darin unterscheiden, welche Steueraufgaben vorgegeben werden, wie genau diese definiert sind und welche kognitiven Anforderungen das Szenario stellt. Daneben ist häufig explizit oder implizit ungeklärt, was damit erfasst werden soll. Weiter wird das dadurch erschwert, dass weder in der Problemlöseforschung noch in der Kybernetik eine klare Definition von Komplexität existiert.

Wie ist nun der Einsatz des computergestützten Szenarios in der Managementdiagnostik zu bewerten? Ähnlich wie das Assessment-Center wird es vor allem im mittleren Management eingesetzt. Darunter fallen teils wissenschaftlich fundierte, in der Mehrzahl ad hoc in der Praxis entwickelte Managementsimulationen, die in der Individualdiagnostik sowie in Gruppen und im Rahmen von Assessment-Centern Verwendung finden. Solche verschiedenen Einsatzvarianten stehen wiederum vor dem Hintergrund unterschiedlicher Intentionen, wie zum Beispiel zur internen oder externen Personalauswahl für bestimmte Positionen oder zur Personalentwicklung, deren Endziel meist auch Personenentscheidungen sind, seien es nun Modifikations-, Klassifikations- oder Selektionsstrategien (Pawlik, 1976).

Problematisch ist zudem, wie die Leistung bei der Bearbeitung des Szenarios ausgewertet wird. Neben direkt aus dem Szenario ermittelten, statusbezogenen Leistungsmaßen, häufig das Gesamtvermögen, stehen prozess-

orientierte Verhaltensmaße zur Verfügung, die über Lösungsstrategien Aufschluss geben sollen und die eigentlich interessante Information darstellen. Die Verhaltensmaße liefern jedoch kaum zuverlässigere Informationen über die Fähigkeiten der Teilnehmer als die statusorientierten Paper-pencil-Tests, sondern eher noch weniger, berücksichtigt man beispielsweise ihre Testgütekriterien wie Reliabilität oder Kriteriumsvalidität, die nicht einheitlich positiv ausfallen (Hasselmann, 1993). Ob auf der anderen Seite die höchst subjektiven und qualitativ verstandenen konstellativen Prozessbetrachtungen gegenüber den Teilnehmern und den Verantwortlichen zu rechtfertigen sind, bleibt ebenfalls fraglich.

Obwohl es inzwischen technisch möglich ist, das Bearbeitungsverhalten am Rechner zu protokollieren und sinnvoll zu aggregieren, gibt es dazu wenig psychologische Fachliteratur. Für die Schneiderwerkstatt wurden Trends bei der Entwicklung des Gesamtvermögens über die Takte erfasst (Kluwe, Schilde, Fischer & Öllerer, 1991). Funkes DISKO (1992) verdichtet die Protokollinformationen zu Bearbeitungsstrategien, um so weitere diagnostische Informationen neben der Steuerleistung zu erhalten. Selbst neuere Publikationen berichten kaum von erfolgreichen Umsetzungen, wie etwa Goldman, Zech, Biswas und Noser (1999), die sich offensichtlich damit schwer tun, die Lösungsstrategie bei Rechenaufgaben technisch abzubilden.

Ein wichtiger Anspruch eines computergestützten Szenarios stammt aus der Tradition der Forschung zum komplexen Problemlösen und besteht darin, dass es eine größere Nähe zu Alltagssituationen aufweist als Verfahren der Intelligenzdiagnostik, von der sich diese Forschungsrichtung abgrenzen will. Die ökologische Validität wird vorwiegend unter dem Aspekt der Realitätsnähe der semantischen Einbettung beziehungsweise der simulierten Situation und Systemzusammenhänge diskutiert. Allerdings wurde dem gegenüber außer in wenigen Studien die Realitätsnähe der Anforderungen als zweiter wichtiger Aspekt nur selten untersucht (Funke, 1992) und auch kaum separat diskutiert.

Eng mit dem Anspruch der Realitätsnähe verbunden ist die Annahme vieler Vertreter des Forschungsparadigmas zum komplexen Problemlösen, dass ihr Szenario eine intellektuelle Leistungskomponente erfasse, welche die mit klassischen Intelligenztests erfassten Fähigkeiten beinhalte und außerdem darüber hinausgehe. Damit verfüge das Konstrukt über inkrementelle Validität. Diese Behauptung wurde seither häufig in Frage gestellt, beispielsweise durch Süß (1996), der in seinen Untersuchungen zeigen konnte, dass sich die Leistung im Szenario Hemdenfabrik vollständig durch Wissens- und Intelligenzmerkmale erklären lässt.

Ob dies an der Operationalisierung des Szenarios liegt, wäre noch zu diskutieren. Auf jeden Fall aber lässt sich daraus gemeinsam mit Süß (1996)

resümieren, dass umgekehrt komplexe Szenarios Wissen und Intelligenz erfassen und die entsprechenden Paper-pencil-Tests ersetzen könnten. Sicher hängt es eher mit dem bereits diskutierten Mangel an detaillierteren Auswertungsparametern zusammen, als mit den Eigenschaften eines Szenarios per se, dass damit „die verschiedenen Fähigkeitsbereiche nicht mehr differenziert erfasst, sondern nur noch als *Amalgam* diagnostiziert werden“ (Seite 205). Sollte darüber hinaus eine Surplus-validity bei anderen Szenarios bestehen, würde das deren Bedeutung für die Managementdiagnostik nur noch erhöhen.

Neben den strukturellen und pragmatischen Aspekten lässt sich das Szenario unter mess- und testtheoretischen Kriterien diskutieren, denen es genügen muss, wenn es zur Eignungsdiagnostik eingesetzt wird. Ideal wäre es, wenn ein Szenario quantitative Skalenwerte mit verschiedener inhaltlicher Bedeutung ausgeben würde, ähnlich wie ein Intelligenz- oder Persönlichkeits-test und damit außerdem dem Existenztheorem (Orth, 1974) genügen würde. Solche Überlegungen wurden in diesem Zusammenhang bisher kaum angestellt oder umgesetzt.

Weiterhin ist wie bei anderen Verfahren die Objektivität zu erörtern. Die Durchführungsobjektivität komplexer Szenarios ist durch den Einsatz von Computern weitgehend gewährleistet, da während der Bearbeitung eine Interaktion zwischen Versuchsleiter und Teilnehmer nicht stattfindet. Eine mögliche Beeinträchtigung besteht in der Dynamik komplexer Szenarios, da sich Maßnahmen nicht linear auf Systemvariablen auswirken und außerdem das Ausmaß der Wirkung stark vom Zeitpunkt abhängt, zu dem diese Maßnahmen getroffen wurden.

Je nach Auswertungsprocedere kann die Auswertungs- und Interpretationsobjektivität eingeschränkt sein, insbesondere, wenn eine qualitative, konfigurale Analyse des Bearbeitungsstils erfolgt, da dies schwer objektivierbar ist. Die Objektivität kann analog zur Konstruktion von Paper-pencil-Tests durch Standardisierung optimiert werden. Dies betrifft die Instruktion, Szenario-Verhalten, rechnergestützte Auswertung, die Bedienoberfläche.

Da interne Reliabilitätsschätzungen wie Chronbach's Alpha oder vergleichbare Kennwerte bei Szenarios bisher nur selten unternommen wurden, ist es meist die Retest-Reliabilität, die anhand der Steuerungsleistung ermittelt wird. Diese liegt bei den meisten Szenarios unterhalb der Minimalanforderung aus der klassischen Testkonstruktion, für die als Faustregel $r = 0.85$ gilt, was aber laut Conrad und Michel (1982) zu hoch gegriffen ist. Weitere Schätzer wären die Stabilität, die Äquivalenz und die interne Konsistenz, indem beispielsweise die Indikatoren für Steuerleistung für jeden Takt bestimmt und korreliert werden. Fraglich ist, ob das aufgrund der gewünschten Systemeigenschaften wie Dynamik und der zu beobachtenden Lerneffekte aus eben diesen Systemvariablen sinnvoll ist.

Worin die Ursachen für diese häufig diskutierten Reliabilitätsmängel liegen, das konnte bis heute nicht geklärt werden, doch es gibt einige Überlegungen dazu. Vermutlich sind sie in typischen Eigenschaften der komplexen Szenarios, wie Dynamik oder Intransparenz zu suchen. Außerdem ist die Abhängigkeit des Gesamtergebnisses von einzelnen zufallsbedingten Entscheidungen höher als bei Paper-pencil-Tests. Dort betreffen Zufallseffekte nur ein einzelnes Item, so dass deren Wahrscheinlichkeit sowie die Bedeutung für das Gesamtergebnis leicht kalkuliert und damit kontrolliert werden kann. Wegen Lerneffekten, die für die Bearbeitung eines Szenarios eigentlich intendiert sind, sind in einem zweiten Testlauf zur Bestimmung der Retest-Reliabilität die Voraussetzungen von Seiten des Probanden nicht mehr mit denen vom ersten Durchlauf vergleichbar.

Außerdem kann die Kriteriumsvalidität als Korrelation mit Berufserfolg berechnet werden, was aufgrund des Aufwandes seltener geschieht. Auch dazu liegen sehr heterogene Befunde vor, die möglicherweise auf die uneinheitliche Operationalisierungen zurückzuführen sind, aber auch auf die kleinen Stichproben der vorliegenden Untersuchungen (Hasselmann, 1993). Berücksichtigt man die inhärenten Probleme vieler computergestützter Szenarios, wie zum Beispiel die häufig diskutierte mäßige Reliabilität selbst unter Laborbedingungen, verwundert es nicht, dass dieses Verfahren in der Praxis eher geringe Kriteriumsvalidität aufweist. Inhaltsvalidität im Sinne einer Korrespondenz von beruflichen Anforderungen und Systemanforderungen wurde insbesondere für den Führungsbereich noch kaum bestimmt.

Damit lässt sich die Frage, ob das computergestützte Szenario für die Managementdiagnostik brauchbar ist, nicht allgemein beantworten. Im Unterschied zu anderen Verfahren ist noch wenig darüber bekannt, in welcher Form das Szenario sinnvoll einzusetzen ist. Neben den beruflichen Anforderungen von Managementaufgaben müssen außerdem die Gestaltungsmöglichkeiten eines computergestützten Szenarios und deren Bedeutung noch exploriert und spezifiziert werden, um eine differenziertere Bewertung vorzunehmen.

2.4.7 Arbeitsprobe

Unter einer Arbeitsprobe versteht man standardisierte Aufgaben, die inhaltlich valide und äquivalente Stichproben des erfolgsrelevanten beruflichen Verhaltens darstellen. Arbeitsproben beruhen auf zwei Grundannahmen. Zum einen soll zukünftiges Verhalten am besten durch vergangenes Verhalten vorhersagbar sein, zum anderen soll der Prädiktor dem Kriterium möglichst ähnlich sein. Dennoch hat der Begriff der Arbeitsprobe mehrere Bedeutungen. Robertson und Kandola (1983) unterscheiden vier Klassen von Arbeitsproben.

Psychomotorische Aufgaben entsprechen der klassischen Herkunft von

Arbeitsproben aus dem Fertigungsbereich und der deutschen Eignungsdiagnostik der zwanziger Jahre (Giese, 1924). Individuelle Entscheidungsprozesse verlangt eine weitere Kategorie von Arbeitsproben, die typischerweise als Postkorbübungen umgesetzt werden. Gruppendiskussionen oder -entscheidungen verlangen ähnliches in Gruppensituationen. Berufsbezogene Wissenstests haben am wenigsten mit dem ursprünglichen Ansatz der Arbeitsprobe zu tun.

Die letzten drei Kategorien wurden für nicht-manuelle Tätigkeiten entwickelt, wie sie auch Führungspositionen darstellen. Außerdem ist die Abgrenzung gegenüber Tests schwierig und uneinheitlich, da häufig Arbeitsproben als Tests bezeichnet werden, sobald sie in standardisierter und normierter Form vorliegen (Schuler, 1996). Dementsprechend findet sich in englischsprachiger Literatur häufig die Bezeichnung *work samples test*. Weiterhin ist für nicht-manuelle Tätigkeiten die Abgrenzung von Arbeitsproben und Assessment-Center-Übungen wie etwa Fallstudien problematisch und uneinheitlich.

Für den Einsatz einer Arbeitsprobe sprechen sowohl eine hohe Akzeptanz bei allen Beteiligten (Robertson & Kandola, 1983), die vor allem auf die Augenscheinvalidität zurückzuführen ist, als auch hohe Validitätskoeffizienten. Die Metaanalyse von Schmitt, Gooding, Noe und Kirsch (1984) berichtet für Arbeitsproben eine durchschnittliche Korrelation mit beruflichen Erfolgskriterien von $r=0.38$, Hunter und Hunter (1984) errechneten einen Koeffizienten von $r=0.54$. Dabei kovariieren diese Validitätswerte mit der Art der Arbeitsprobe und der Art der Kriterien. Für psychomotorische Arbeitsproben sowie Postkorb- und Gruppenübungen resultieren dann höhere Werte, wenn sie mit Vorgesetztenbeurteilungen korreliert werden, während berufsbezogene Wissenstests höher mit dem Ausbildungserfolg korrelieren (Robertson & Kandola, 1983).

Gleichzeitig sind mit der Arbeitsprobe einige Probleme verbunden. Typischerweise werden sie auf der Basis einer Tätigkeitsanalyse der in Frage stehenden Position entwickelt. Dies bedeutet einen hohen Konstruktionsaufwand sowie eine geringe Generalisierbarkeit auf andere Tätigkeiten. Vor allem, wenn die Tätigkeit sehr vielfältig oder abstrakt ist und Interaktionen mit anderen Personen beinhaltet, ist es schwieriger, Arbeitsproben zu generieren als für industrielle Routinearbeiten im Produktionsbereich. Damit ist die Arbeitsprobe für die Managementdiagnostik problematischer als für die Eignungsdiagnostik bei Arbeitern in der Fertigung.

Weitere Kritikpunkte (Barrett, 1992) richten sich auf die Stabilität der Validitätswerte, da diese über einen längeren Prognosezeitraum abnehmen, insbesondere, wenn sich die Arbeitstätigkeit ändert, was häufig vorkommt. Außerdem differenzieren Arbeitsproben nicht nach den enthaltenen Teilfä-

higkeiten in der Arbeitsstichprobe, was die Übertragbarkeit und Stabilität ebenso beeinträchtigt wie die Möglichkeit, die Arbeitsprobe an veränderte Gegebenheiten anzupassen. Damit hat die Arbeitsprobe je nach ihrer Gestaltungsvariante das Potenzial, beruflichen Erfolg vorherzusagen, das aber nicht in allen Fällen optimal ausgeschöpft wird.

2.4.8 Biografische Daten

Die Auswertung der Bewerbungsunterlagen stellt in der Regel den ersten Schritt bei der Auswahl von Mitarbeitern dar. In dieser Phase können sie im Falle einer Ablehnung bereits entscheidend sein. Daneben dienen sie meist als Grundlage für weitere Gespräche und Verfahren. Schließlich werden sie in die abschließende Entscheidung einbezogen. Damit ist ihr Stellenwert insgesamt nicht höher als derjenige der anderen eignungsdiagnostischen Verfahren. Für die optimale Verwertung der Unterlagen gibt es bisher wenig Befunde oder Modelle. In der Praxis werden üblicherweise verschiedene Aspekte der Unterlagen betrachtet (Schuler, 1996).

Zunächst wird die Bewerbungsmappe nach formalen Gesichtspunkten bewertet, wie weit die Präsentation den Standards der ausgeschriebenen Position entspricht. Außerdem prüft man, ob die erforderliche Ausbildung anhand von Zeugnissen und Bescheinigungen vorliegt sowie Zusatzqualifikationen. Häufig werden diese Belege mit den zeitlichen Angaben im Lebenslauf verglichen. Schulnoten und Studienleistungen werden zur Prognose von weiteren Ausbildungsleistungen genutzt, sind aber zur Vorhersage von Berufserfolg weniger brauchbar. Arbeitszeugnisse und Referenzen sind aufgrund fehlender Standards weniger verlässlich, wobei mündliche Referenzen meist aussagekräftiger sind als schriftliche. Ergänzende anforderungsspezifische Aspekte wie Berufserfahrung oder Mobilität werden je nach Art der Position gezielt abgeprüft.

Untersuchungen zur prognostischen Validität der Bewerbungsunterlagen wurden insbesondere für die Einzelinformationen durchgeführt, die aus den Unterlagen hervorgehen, wie Schulnoten oder Studienleistungen (Metaanalyse von Baron-Boldt, Funke & Schuler, 1989), aber kaum in der Gesamtheit oder bezogen auf Urteilsmodelle. Die soziale Validität oder Akzeptanz der Bewerbungsunterlagen ist ebenfalls wenig bekannt. Insofern blieb bisher offen, welchen Beitrag eine systematische Auswertung der Bewerbungsunterlagen als Prädiktor zur Aufklärung oder Vorhersage des Kriteriums des beruflichen Erfolgs leisten kann, und wie dieses Procedere unter Gesichtspunkten der Objektivität und Reliabilität zu bewerten ist.

Ebenfalls hauptsächlich zur Personalauswahl wird der biografische Fragebogen eingesetzt, der, einem systematisierten Prinzip folgend, aus dem

biografischen Hintergrund einer Person und ihrem Verhalten in der Vergangenheit das zukünftige Verhalten sowie den Berufserfolg vorhersagen soll. Dabei gewichtet man jedes Item nach seinem Beitrag zur Erfolgsprognose an der gegebenen Stichprobe. Das erhöht die Validität für die gegebene Anwendungssituation, reduziert jedoch gleichzeitig die Generalisierbarkeit und ist mit zusätzlichem Aufwand verbunden.

Neben diesem Prinzip der empirischen Validierung unterliegen die Items biografischer Fragebogen weiteren Konstruktionsprinzipien (Schuler, 1996). Die Fragen können sich dabei auf vergangenes oder zukünftiges Verhalten richten sowie auf tatsächliches oder hypothetisches. Inhaltlich sind sie entweder konkret oder allgemein, in der Regel beziehen sie sich auf nachprüfbare Sachverhalte, die faktisch oder interpretativ abgefragt werden. Dabei können außerdem äußere Ereignisse oder innere Vorgänge angesprochen werden. Schließlich beziehen sich die Fragen auf Leistungen, Aktivitäten oder Anforderungen, die mit persönlichen oder beruflichen Erfahrungen verbunden sind. Knoblauch (1990) grenzt biografische Items von Persönlichkeitsitems, Interessenitems, Meinungs- und Wissensfragen ab, die sich prinzipiell auf die gleichen Inhalte beziehen können, wobei nur biografische Items diese Inhalte mit konkreten Erfahrungen in Beziehung setzen.

Zur Kriteriumsvalidität von biografischen Fragebogen gibt es sehr viele Untersuchungen und etliche Metaanalysen, die insgesamt für eine recht hohe Validität sprechen, womit der biografische Fragebogen die höchste Validität der hier vorgestellten Verfahren hat. Dabei hängt die Validität ähnlich wie bei anderen Instrumenten von der Art des Kriteriums ab. Die Gehaltshöhe als Kriterium konnte durchschnittlich mit $r=0.52$ vorhergesagt werden, während der Zusammenhang mit der Ausbildungsleistung nur $r=0.23$ beträgt (Schmitt, Gooding, Noe & Kirsch, 1984).

Andere Metaanalysen kommen auf einen Validitätskoeffizienten von $r=0.37$ (Hunter & Hunter, 1984). Neben der Art des Kriteriums ist der Erhebungszeitpunkt relevant, wobei der Zusammenhang mit dem beruflichen Erfolgskriterium über die Zeit geringer wird. Das erfordert unter Umständen eine erneute Gewichtung der biografischen Items bezogen auf das später erhobene Kriterium. Weiterhin hängt die Höhe des Zusammenhangs von der Zielgruppe ab. Am häufigsten wird der biografische Fragebogen bisher an Verkaufspersonal im Versicherungsaußendienst und an Führungskräften eingesetzt.

Im Zusammenhang mit der Konstruktvalidität werden in der Literatur verschiedene Varianten des biografischen Fragebogens diskutiert. Generell zeichnet sich das Verfahren durch eine konstruktthetogene Struktur aus, wodurch es sich von den faktoriell homogenen Skalen klassischer Persönlichkeitstests unterscheidet. Worin die abgebildeten Konstrukte bestehen,

ist umstritten. Der Fragebogen kann hinsichtlich der beruflichen Anforderungen frühere Leistungen erfragen, was die kriterienbezogene Validität des Verfahrens steigert. Alternativ könnte man mit biografischen Fragebogen psychologische Konstrukte aus dem Leistungs-, Einstellungs- und Interessenbereich umsetzen. Die Konstruktvalidität ist in diesem Zusammenhang ohnehin schwierig zu realisieren und zu prüfen, da es kaum Theorien biografischer Daten gibt, obwohl einige Ansätze denkbar sind (Jäger, 1992). Zugunsten der Validität und Generalisierbarkeit sollten Konstrukt- und Anforderungsorientierung vorliegen.

2.5 Fazit

Es wurde versucht, ein allgemeines Modell der Evaluationsforschung auf die Managementdiagnostik zu übertragen, und für die sechs Schwerpunkte skizzenhafte Anwendungsmöglichkeiten aufzuzeigen. Zumindest berücksichtigt dieses Modell wesentlich mehr Aspekte der Managementdiagnostik, als dies für eine reine Kriteriumsvalidierung der Fall ist. Anschaulicher soll dies in der eigenen Umsetzung mit Evaluationsstudien aus verschiedenen Perspektiven werden, die alle Beiträge zur Managementdiagnostik liefern beziehungsweise dort stattfanden.

Das „Treatment“ in der Managementdiagnostik besteht in den diagnostischen Instrumenten. Die wichtigsten wurden in ihrer Brauchbarkeit hier diskutiert. Jedoch kann daraus nicht eindeutig abgeleitet werden, welche davon sich besonders eignen. So hat jedes dieser Verfahren, wie etwa das Assessment-Center so viele methodisch und inhaltlich verschiedene Durchführungsvarianten, dass diese nicht einheitlich beurteilt werden können. Vielmehr hängt ihre Brauchbarkeit davon ab, wie weit sie methodischen Standards genügen und wie weit sie in einem inhaltlichen Zusammenhang mit dem Anwendungszweck stehen. Gleichzeitig ist es wichtig, Bewertungsstrategien zu erarbeiten, die stärker als die bekannten Validierungsstrategien auch die Praxissituation berücksichtigen, insbesondere, wenn in der Anwendungssituation eine Evaluation der eignungsdiagnostischen Maßnahme stattfinden soll.

Kapitel 3

Führungsforschung

3.1 Übersicht

Chen (1990) weist bei der Darstellung seines Evaluationsmodells darauf hin, dass Evaluation nicht unabhängig von inhaltlichen Theorien stattfinden kann, sondern bei jeder Maßnahme auch wissenschaftliche Ansätze und Befunde zum entsprechenden Phänomen genutzt werden sollten. Insbesondere für eine kausative Evaluation intervenierender Mechanismen ist eine gute Kenntnis der Forschungslage nötig, um die wichtigen Prozesse und Variablen, die im jeweiligen Anwendungsfeld eine Rolle spielen, möglichst gut zu präzisieren. Deshalb soll hier eine Bestandsaufnahme vorgenommen werden, welche Ansätze und Methoden die Führungsforschung beinhaltet, die das Thema meist unabhängig von den psychometrischen Problemen der Managementdiagnostik behandelt. Dennoch gibt es Schnittstellen, wie etwa die Methoden der Führungsforschung, oder die Traittheorien der Führung und charismatische beziehungsweise transformationale Führungsmodelle.

Zum Thema Führung oder Management gibt es seit Anfang des zwanzigsten Jahrhunderts eine große Menge an Literatur (Borman & Brush, 1993). Sie unterscheidet sich in ihrer Zielsetzung, in den forschungsstrategischen Hintergründen und in den empirischen Grundlagen, da dieses Gebiet einen Anwendungsschwerpunkt verschiedener sozialwissenschaftlicher Disziplinen darstellt (Van Fleet, 1975). Die Zielsetzung der Managementliteratur besteht nicht allein darin, Erkenntnisse für Wissenschaft oder Praxis zu gewinnen, sondern oft gibt sie auch Verhaltensempfehlungen für Führungskräfte oder deren Mitarbeiter. Die Untersuchungen der Führungsforschung haben jedoch alle das Ziel, mit ihren Befunden Führungserfolg zu erklären, vorherzusagen, oder zu beeinflussen. Die Übersicht verschiedener Führungstheorien in diesem Kapitel stellt die jeweiligen Ansätze und Befunde verschiedener Diszi-

plinen problemspezifisch dar, das heißt welche Befunde auf diagnostizierbare Anforderungen an Führungskräfte schließen lassen. Umfassendere Darstellungen zur Führungsforschung enthalten zum Beispiel die Bücher von Yukl (2002), Rosenstiel (1994) oder Neuberger (2002).

3.2 Forschungsstrategien

Die Ergebnisse wissenschaftlicher Führungstheorien hängen auch von den empirischen Methoden ab, mit denen sie entwickelt oder überprüft werden, sowie von der Perspektive, unter der die Führungsforschung betrieben wird. Deshalb könnte man die Zugänge nach methodischen Aspekten gliedern, das heißt nach den verwendeten Erhebungsverfahren oder nach inhaltlichen Aspekten, also den Perspektiven, unter denen die Erhebung mit verschiedenen Methoden stattfindet. Im Folgenden erfolgt die Darstellung nach inhaltlichen Gesichtspunkten gegliedert.

In der deutschsprachigen Literatur gibt es weniger Material zum Verhalten von Führungskräften als in der angloamerikanischen Fachliteratur. Das ist wohl mit ein Grund, dass die Managementdiagnostik im deutschen Raum auf weniger Informationen über die Inhalte und Anforderungen in Führungsaufgaben zurückgreifen kann. Vielmehr muss sie sich auf korrelative Befunde über den Zusammenhang zwischen den Ergebnissen von einzelnen Diagnoseinstrumenten auf der einen Seite und allgemeinen beruflichen Bewährungskriterien auf der anderen Seite verlassen. Dies kann zwar die Relevanz einer gezielten Leistungsdiagnostik im Nachhinein untermauern, aber im Vorfeld keine strukturierenden und theoretisch begründeten Hilfen für eine maßgeschneiderte Personaleignungsdiagnostik bieten.

3.2.1 Definition

Zunächst ist aber eine Arbeitsdefinition des Begriffs Führung festzulegen, da zum Beispiel Stogdill (1974) angesichts der Führungsliteratur konstatierte: „There are almost as many definitions of leadership as there are persons who have attempted to define the concept.“. An dieser Situation hat sich seither wenig geändert. In der angloamerikanischen Literatur gibt es viele Definitionen von Führung, auf englisch *leadership*, die teilweise den Definitionen von *management* entsprechen, nur Überschneidungen aufweisen oder kontrastiert werden (Zaleznik, 1977). Zur Verwirrung trägt bei, dass in der deutschsprachigen Literatur gelegentlich den Begriffen Management und Führung als drittes Konzept *leadership* mit eigenständiger Bedeutung und zusätzlichen Wertungen gegenüber gestellt wird.

Weiterhin wird eine strukturelle oder positionale Auffassung im Unterschied zu einem emergenten oder dynamischen Verständnis von Führung diskutiert. Führung im strukturellen Verständnis bedeutet, dass eine Person eine übergeordnete Machtposition inne hat, deren Ausübung durch formale Regelungen und präzise Rollenerwartung bestimmt ist (Kerr & Jermier, 1978). Diese Variante entspricht der situativen Führung, wie sie Neuburger (2002) versteht. Dagegen liegt aus der Sicht der dynamischen Auffassung Führung nicht in äußeren Strukturen, sondern in der Persönlichkeit des Führenden begründet, und bildet sich weniger strukturierten Situationen heraus, indem bestimmte Personen aufgrund ihrer besonderen Eigenschaften als Führer anerkannt werden (Hofstätter, 1986). Strittig ist außerdem, ob Führung im Unterschied zu Management nur auf soziale Situationen begrenzt sein soll und ob eine solche Auffassung nicht generell als unnötige Einengung anzusehen ist (Yukl & Van Fleet, 1992).

Die vorliegende Arbeit will jedoch keinen weiteren Beitrag zu diesen seit langem geführten Kontroversen liefern, sondern stützt sich auf eine weiter gefasste Arbeitsdefinition von Yukl (1994). Demzufolge wird Führung oder Management in gleicher Bedeutung verwendet und ist ein Prozess, innerhalb dessen auf verschiedenen Ebenen eine Einflussnahme stattfindet. Dabei kann der Einfluss auf Aufgabenzielen und Strategien einer Gruppe oder Organisation ausgeübt werden, oder auf Personen in der Organisation, um die Strategien umzusetzen und die Ziele zu erreichen, oder zuletzt darauf, dass Gruppenstrukturen und Organisationskultur identifiziert und aufrecht erhalten werden. Dies deckt sich im wesentlichen mit der Basisdefinition von Management aus der betriebswirtschaftlichen Literatur (Wehrich & Koontz, 1993), wonach Management ein Prozess ist, der eine Umgebung gestaltet und aufrechterhält, in der Personen, die zusammenarbeiten, bestimmte Ziele effizient erreichen.

Damit wurde eine bewusst allgemein gehaltene Führungsdefinition gewählt. Sie berücksichtigt keine weitere Facetten von Führung, die sich im Alltag stark unterscheiden kann. So ist zum Beispiel wichtig, ob der Schwerpunkt einer Führungskraft eher auf Fach- oder auf Führungsaufgaben liegt, aus welchem Bereich die fachlichen Aufgaben kommen, in welcher Führungsebene sich die Führungskraft befindet und damit eher strategische oder operative Verantwortung hat. Ebenso spielen weitere Situationsfaktoren eine Rolle, wie die Branche, die Kultur oder die Unternehmensgröße.

Außerdem kommt dazu häufig die tatsächliche informelle Verteilung von fachlicher und administrativer Verantwortung, die in einer Abteilung unabhängig von den offiziellen Stellenbeschreibungen ausgehandelt wird, und von den persönlichen Kompetenzen der Beteiligten abhängt (Badke-Schaub, 2001). Damit werden häufig Teile von Führungsfunktionen nicht mehr von

der offiziell dazu bestimmten Person ausgeübt, sondern von den Mitarbeitern übernommen, während sich umgekehrt unter bestimmten Umständen die Führungskraft stärker als vorgesehen Fachaufgaben widmet.

3.2.2 Funktionale Analysen

Analytisch-funktionale Beschreibungen des Managementprozesses stellen den historisch ältesten Zugang zu dieser Problematik dar. Ihr Ziel ist es, die Arbeit von Führungskräften als Ausübung von Funktionen zu rekonstruieren, wobei Management eine spezifische Funktion ist, die in arbeitsteiligen Organisationen erforderlich ist. Sie führt die Steuerungsprozesse innerhalb und zwischen Organisationen aus, um das erfolgreiche Weiterbestehen der Organisation als System zu gewährleisten. Damit sollten Manager ähnliche Funktionen erfüllen wie andere organisationale Prozesse. Somit werden Organisationen implizit oder explizit aus einer systemtheoretischen Sichtweise beschrieben und als übergeordnete Funktionseinheiten gesehen, die sich aus untergeordneten Organisationseinheiten zusammensetzen, von denen eines die Administration ist, welche wiederum die Managementfunktionen beinhaltet.

Nach Fayol (1929) sind zunächst insgesamt sechs Grundvorgänge für das Funktionieren eines Unternehmens notwendig und typisch. Technische Vorgänge decken die Produktion ab, kommerzielle Vorgänge beinhalten Aktivitäten wie Einkauf und Verkauf, finanzwirtschaftliche Vorgänge dienen der Kapitalbeschaffung und -verwendung, Sicherheitsmaßnahmen dem Schutz von Vermögen und Personen. Rechnungslegung als fünfte Dimension zielt vor allem auf Inventur, Bilanzen und Statistik. Die sechste Dimension der Administration nimmt bei Fayol eine Sonderstellung ein, da sie als Querschnittsfunktion die anderen fünf sogenannten Sachfunktionen integriert.

Administration umfasst wiederum fünf Dimensionen. Das Begriffspaar Vorausschau und Planung enthält einerseits, künftiges Geschehen gedanklich vorherzusehen, andererseits, zukünftige Handlungen geistig vorwegzunehmen. Beides soll miteinander sowie mit den Unternehmenszielen, realisierten Plänen und humanen, technischen und materiellen Ressourcen in Einklang stehen. Organisation bedeutet, dass materielle und personelle Ressourcen zu Organisationseinheiten zugeordnet und die Beziehung dieser Einheiten untereinander festgelegt werden. Das umfasst zum Beispiel Kontrollspannen, Stellenbesetzung, Leistungsbeurteilung und die Aus- und Weiterbildung.

Mit der Leitung oder Auftragserteilung sind Aktivitäten gemeint, die betriebliche Tätigkeiten im Interesse des Unternehmens veranlassen. Dafür nennt Fayol (1929) einige Voraussetzungen, wie beispielsweise gute Kenntnis des Personals, Ausüben von Vorbildfunktionen, Durchführung von Mitarbei-

terbesprechungen und die Möglichkeit unfähiger Mitarbeiter auszuschneiden. Die Koordination als vierte Dimension stellt sicher, dass die sechs Grundvorgänge des arbeitsteilig organisierten Unternehmens einander zuarbeiten. Koordinationsaufgaben sichern den Informationsfluss und die Kommunikation zwischen Abteilungen, außerdem richten sie die Mitarbeiterinteressen auf das allgemeine Unternehmensinteresse aus. Die letzte Dimension, die Kontrolle, überprüft, ob Aktionspläne in allen Funktionsbereichen umgesetzt wurden, was in Form der Abweichungsanalyse geschieht, und passt die Konzeption von Maßnahmen diesen Plänen an.

Dabei folgen diese fünf Dimensionen in der dargestellten Reihenfolge aufeinander, um sich in einem zyklischen Prozess zu wiederholen. Dies gilt ebenso für alle weiteren funktionalen Ansätze, die manche von Fayol (1929) postulierten Dimensionen als überflüssig weglassen (Beyer, 1970; Wehrich & Koontz, 1993), oder um weitere Dimensionen ergänzen, wie es beispielsweise bei der unter dem Akronym POSDCORB bekannten Klassifikation von Gulick (1937) der Fall ist, die großen Einfluss auf die weitere Forschung im Rahmen des funktionalen Ansatzes hatte.

Innerhalb dieser Forschungsrichtung erfolgt die Datenerhebung meist in Form schriftlicher Befragungen mit strukturierten und weitgehend standardisierten Fragebögen, welche die Häufigkeit und Wichtigkeit von Verhaltensweisen einschätzen lassen, die den jeweiligen Dimensionen zuzuordnen sind. Dabei stützt sich die Auswertung überwiegend auf Häufigkeitsverteilungen, um den Anteil der Arbeitszeit nach Funktionen zu ermitteln. Manchmal werden auch strukturentdeckende Verfahren, in der Regel Faktorenanalysen, durchgeführt, um die Klassifikationen zu modifizieren. Die Mehrzahl der Studien arbeitet mit vorgegebenen Kategorisierungen aus den gewählten Ansätzen der Führungsliteratur.

Schirmer (1992) gibt eine Übersicht über empirische Untersuchungen des Arbeitsverhaltens von Managern, die auf funktionsanalytischen Konzepten aufbauen. Diese Studien unterteilt er nach ihrer Aussageeinheit. Entweder schließen sie dabei deckungsgleich mit der Untersuchungseinheit auf das Verhalten als erste Aussageeinheit, oder die Ergebnisse sollen Positionen personenunabhängig beschreiben.

Weiter gliedern sich die funktionsanalytischen Studien ähnlich wie multivariate statistische Analysen nach der Art des Verdichtungs-niveaus, das auf die Merkmalsebene oder auf die Objektebene zielt. Die Merkmalsebene fasst das Arbeitsverhalten zu Beschreibungsdimensionen zusammen, die Objektebene enthält mehrere Objekte mit ähnlich ausgeprägten Merkmalen und ermöglicht die Identifikation von Jobtypen. Damit legt Schirmer (1992) ein Vierfelder-Schema fest, in das er die funktionsanalytischen Untersuchungen einordnet.

Verhaltenszentrierte Studien auf der Merkmalsebene erfassen mehrere verschiedene Verhaltensweisen, die gleichzeitig bewertet werden. Daraus können nicht nur Verhaltensmuster im Sinne eines Durchschnittsprofils gewonnen, sondern auch Aussagen über ihre Effizienz getroffen werden, die allerdings subjektive Einschätzungen darstellen. Die standardisierte Erfassung begrenzt außerdem die Beschreibung auf die vorgegebenen Aspekte. Verhaltenszentrierte Ansätze auf der Objektebene ermöglichen die Bildung von Jobtypen, die auf den Funktionen unterschiedliche Schwerpunktverteilungen haben. Gleichzeitig kann ein minimales gemeinsames Tätigkeitsprofil festgestellt werden.

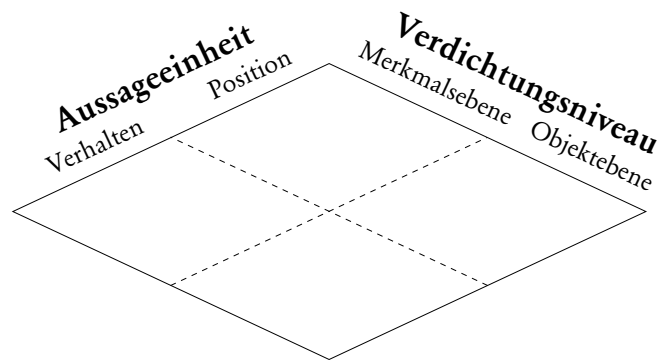


Abbildung 3.1: Klassifikationssystematik für Aktivitätsstudien anhand funktionaler Kategorien (nach Schirmer, 1992).

Das stützt die Annahme von Fayol (1929), dass Managementfunktionen Querschnittcharakter haben. Positionszentrierte Untersuchungen auf der Merkmalsebene basieren auf einem Rollenkonzept, wobei die Rollen die klassischen Grundfunktionen enthalten. Sie repräsentieren Verhaltenserwartungen und Verantwortlichkeiten, die mit einer Position verknüpft sind. Daraus resultieren zugleich standardisierte und differenzierte Beschreibungskonzepte, die in verschiedenen Anwendungsfällen einzusetzen sind, und außerdem für personalwirtschaftliche Aufgabenstellungen wie zum Beispiel zur Arbeitsbewertung genutzt werden können.

Auf der Objektebene werden Dimensionen zur Positionsbeschreibung weiter verdichtet, um Typen von Managementpositionen zu identifizieren. Bei diesen Untersuchungen wurden sehr heterogene Analysesysteme verwendet, weshalb die Befunde schwer zu vergleichen oder zu verallgemeinern sind. Dennoch ermöglichen sie die Klassifikation verschiedenster Managementdimensionen, wobei der Abstraktionsgrad der Studien schwankt.

Insgesamt resultieren aus den verschiedenen funktional orientierten Analysen, die bis heute in dieser Form durchgeführt werden, sehr stabile Ergebnisse über die Kernfunktionen von Managern, die recht nah an dem Ausgangsmodell von Fayol (1929) liegen. Darin besteht gleichzeitig die Stärke und die Schwäche des Ansatzes, der so durch viele ähnliche Befunde gestützt wird. Gleichzeitig wird durch die Art der Erhebung die Bandbreite möglicher Beobachtungen, die nicht in das Raster passen, stark eingeschränkt.

Mit diesem Instrumentarium ist es demzufolge schwierig, die zugrundeliegenden Modelle zu erweitern, ebenso wie andere explorative Erhebungsstrategien sich kaum anwenden lassen. Auch für eine theoriegeleitete Analyse des Funktionswandels ist es nur bedingt nutzbar. Vorteilhaft ist dagegen, dass das Verhalten von Führungskräften in einen größeren Gesamtzusammenhang eingeordnet wird und damit eine Makroperspektive anbietet, die den anderen Ansätzen fehlt, die dafür im Detail konkreter und präziser sind.

3.2.3 Anforderungsanalysen

Berufliche Anforderungen sind Merkmale des Arbeitsplatzes, werden aber als persönliche Leistungsvoraussetzungen dargestellt, die der Stelleninhaber für die erfolgreiche Bewältigung seiner Arbeitsaufgaben mitbringen muss (Hacker, 1983). Dies umfasst Eigenschaftsanforderungen, Qualifikationsanforderungen, Verhaltensanforderungen und Ergebnisanforderungen. Eigenschaftsanforderungen sind strukturelle Merkmale einer Person, insbesondere Fähigkeiten, Persönlichkeitsmerkmale und Interessen. Unter Qualifikationsanforderungen versteht man erworbene Kenntnisse und Fertigkeiten. Verhaltensanforderungen beschreiben das notwendige Verhaltensrepertoire sowie kognitive Prozesse, und Ergebnisanforderungen beziehen sich auf die geforderten Lösungen und Arbeitsergebnisse.

Anforderungen können aus einer Arbeitsanalyse abgeleitet werden. Darunter versteht man eine Gruppe von Verfahren zur Sammlung und Analyse arbeitsbezogener Informationen, die sich hinsichtlich Anwendungszweck, Inhalt, theoretischer Fundierung und formaler Eigenschaften unterscheiden können. Ihre Anwendungsschwerpunkte liegen in der Eignungsdiagnostik und der Arbeitsplatzgestaltung (Schuler, 1989). In der Eignungsdiagnostik soll die Arbeitsanalyse die Verbindung zwischen Situationsdiagnostik und Personendiagnostik herstellen, indem aus ihr diagnostizierbare Personeneigenschaften abgeleitet werden.

Um nun eine Anforderungsanalyse für Führungskräfte durchzuführen, müssten die Elemente der Arbeitstätigkeit zunächst empirisch erfasst werden. Dieser deskriptive Vorgang besteht entweder zunächst in einer möglichst erschöpfenden explorativen Beobachtung und Aufzeichnung des Arbeitspro-

zesses, um dieses Material später zu kategorisieren und zu abstrahieren. Anschließend oder alternativ dazu werden die Beobachtungen in ein theoretisch abgeleitetes oder empirisch ermitteltes Kategoriensystem eingeordnet. Ideal für die intendierte Verwendung der Ergebnisse wäre eine eigenschaftsorientierte Anforderungsanalyse. Laut Kannheiser (1995) „ist jedoch festzustellen, dass Anforderungsanalysen für konkrete Management- und Führungspositionen nur sporadisch durchgeführt wurden“.

Abgesehen von wenigen Ausnahmen wie etwa bei Fleishman, Mumford, Zaccaro, Levin, Korotkin und Hein (1991), steht dabei eine stark personalistische Betrachtungsweise im Vordergrund. Das bedeutet, die managementdiagnostische Praxis orientiert sich meist situationsunabhängig an globalen Verhaltens- oder Persönlichkeitsmerkmalen, die aus Annahmen und Theorien abgeleitet werden, deren Gültigkeit man nicht kennt. Die wenigen arbeitsanalytischen Vorgehensweisen, die dokumentiert sind, beispielsweise bei Hemphill (1959), weisen teilweise erhebliche methodische Mängel auf, die in der fehlenden Replizierbarkeit der faktorenanalytisch ermittelten Skalenstrukturen oder in der mangelnden Differenzierungsfähigkeit der Anforderungen liegen.

Nicht nur in der Managementdiagnostik, sondern generell in der Personaleignungsdiagnostik, wird die Arbeits- oder Anforderungsanalyse häufig als akademisches, nicht praxisrelevantes Vorgehen angesehen, das vielleicht sinnvoll wäre, aber viel zu unökonomisch im Verhältnis zum erwarteten Validitätszuwachs ist. Weiterhin wird kritisiert, dass bei Aufgaben mit hohem Handlungsspielraum oder hoher Änderungsgeschwindigkeit bezüglich der Aufgaben Arbeitsanalysen nicht realisierbar seien, und solche als besonders relevant angesehenen Anforderungen wie soziale Kompetenz nicht erfassen (Schuler & Funke 1993). Dabei stellt sich jedoch im nächsten Schritt die Frage, woher man sonst wissen will, dass dies eine relevante Anforderung ist.

Solange die Frage des Nutzens von Anforderungsanalysen aber nicht empirisch überprüft wird, führt auch diese Diskussion nicht weiter, ganz gleich, mit welchen Argumenten sie geführt wird. Vielmehr besteht ohne Anforderungsanalyse „die Gefahr, falsche Vorstellungen über die Managementtätigkeit zu perpetuieren“ (Kannheiser, 1995). Dagegen konnten Metaanalysen zeigen, dass strukturierte Interviews und Persönlichkeitstests höhere Validitätskoeffizienten erreichen, wenn sie auf Anforderungsanalysen aufbauen, wie Cook (1998) berichtet. Beide Verfahren erreichten in Verbindung mit Anforderungsanalysen die höchsten dafür bekannten Validitätskoeffizienten.

Generell ist damit die Anforderungs- oder Tätigkeitsanalyse in der Form, wie sie aus der Arbeitspsychologie bekannt ist, in der Führungsforschung ein untypischer Zugang, und wenn sie durchgeführt wird, dann genügt sie häufig

nicht wissenschaftlichen Anforderungen (Rosenstiel, 2001). Da aber gerade die Tätigkeit einer Führungskraft nicht nur praktische Relevanz etwa für ein Unternehmen hat, sondern sich außerdem wesentlich mehr wissenschaftliche Publikationen damit beschäftigen als mit anderen beruflichen Tätigkeiten, erscheint es lohnenswert, die Anforderungsanalyse auch auf die Führung zu adaptieren.

3.2.4 Aktivitätsstudien

Die ersten Aktivitätsstudien im amerikanischen Sprachraum wurden in den fünfziger Jahren durchgeführt und wollten möglichst konkret beschreiben, was Manager alltäglich tun. Dies stellte eine Reaktion auf die bisherigen analytisch-funktionalen Modelle dar. Um die Datenbasis zu erweitern, soll eine möglichst genaue und umfassende Dokumentation des Verhaltensstroms im alltäglichen Arbeitsvollzug erfolgen, um damit eine breitere empirische Basis für die Theorienbildung zu erhalten. Von Aktivitätsbeschreibungen werden wichtige Anregungen für die Lösung praktischer Probleme wie Training und Auswahl erwartet (Stewart, 1976), wofür funktionale Analysen ungeeignet seien.

Dazu werden hauptsächlich Verfahren der Selbst- oder Fremdbeobachtung eingesetzt, sowie mündliche oder schriftliche Befragung. Bei der Selbstbeobachtung sollen Manager alle täglichen Aktivitäten in standardisierte Protokollbögen, so genannte Tagebuchblätter eintragen. Bei der fortlaufenden unmittelbaren Fremdbeobachtung begleitet ein Beobachter eine Führungskraft den Tag über, der alle Aktivitäten strukturiert oder unstrukturiert festhält. Diese Methoden kamen zunächst einzeln, später auch kombiniert zum Einsatz, um eine facettenreichere Beschreibung des Arbeitsverhaltens zu ermöglichen und gleichzeitig die speziellen Schwächen der einzelnen Erhebungsmethoden zu kompensieren. Die Auswertung erfolgt in der Regel mit qualitativen und quantitativ-deskriptiven Verfahren, nur selten mit multivariaten statistischen Analysen.

Ähnlich wie bei den funktionalen Analysen lassen sich die Aktivitätsstudien nach ihren Untersuchungsschwerpunkten einordnen. Schirmer (1992) unterscheidet statisch-analytisch und prozessual-integrativ ausgerichtete Studien. Die Mehrzahl der Untersuchungen ist statisch-analytisch ausgerichtet. Sie wollen das Verhalten von Führungskräften bei der Erfassung in einzelne additive Elemente zerlegen.

Dabei gibt es zwei grundlegende Fragestellungen. Zum einen geht es darum, die Häufigkeit, Dauer und äußeren Umstände der Aktivitäten festzuhalten, zum anderen interessiert, zu welchem Zweck Führungskräfte arbeiten. Integrativ-funktional orientierte Studien sind dadurch gekennzeichnet, dass

sie Zusammenhänge zwischen den beobachtbaren Aktivitäten sowie Wechselbeziehungen zwischen Arbeitsverhalten und organisationalem Kontext in ihre Beschreibungs- und Erklärungsansätze einbeziehen. Dafür wird zusätzlich das nicht äußerlich beobachtbare Verhalten als relevant angesehen.

Im Unterschied zu funktionalen Analysen nehmen die Aktivitätsstudien eine Perspektive ein, die nahe am Akteur liegt. Dadurch können sie sehr konkrete und alltagsnahe Befunde liefern, gleichzeitig verlieren sie die Makroperspektive im Sinne übergeordneter betrieblicher Zusammenhänge aus dem Blick. Dennoch liefert eine Aggregation dieser verhaltensorientierten Beobachtungsdaten ähnliche Resultate wie die funktionalen Ansätze, so dass diese im Unterschied zur ursprünglichen Absicht, die funktionalen Ansätze in das Reich der Mythen zu verweisen, diese vielmehr validieren (Wren, 1990).

Ein bisher ungelöstes Problem besteht in der differenziellen Wirkung von Management-Verhalten, über das bisher nur wenige Befunde vorliegen. Das liegt unter anderem darin begründet, dass die Indikatoren für Effektivität und Effizienz nicht über quantifizierbare Input-Output-Beziehungen darstellbar sind, wie es im Produktionsbereich möglich ist. Ein weiterer, häufig genannter Kritikpunkt an diesen Studien besteht darin, dass diese Art der Erfassung von Managerverhalten der organisationalen und kognitiven Dynamik nicht gerecht würde (Starbuck, 1983). Warum die Methode der Erfassung jedoch strukturelle Ähnlichkeit mit dem erfassten Phänomen haben muss, wird nicht logisch oder auf anderem Weg begründet.

3.2.5 Kategoriensysteme und Taxonomien

Nicht obwohl, sondern weil Kategoriensysteme im Zusammenhang mit allen bisher beschriebenen methodischen Zugängen entwickelt und eingesetzt werden, soll ihnen ein eigenes Teilkapitel gewidmet werden. Kategoriensysteme ermöglichen die Beobachtung von Verhalten, indem sie mögliche Verhaltensweisen benennen und klassifizieren. Sie werden in der Regel im Zusammenhang mit einem bestimmten theoretischen und methodischen Ansatz der Führungsforschung entwickelt, entweder um Beobachtungen in diesem Rahmen zu strukturieren, oder sie sollen auf dem theoretischen Hintergrund Diagnosen des Führungsverhaltens ermöglichen.

Damit müssen solche Kategoriensysteme den bekannten diagnostischen Gütekriterien genügen, auf die das jeweilige Verfahren explizit geprüft werden kann, wie zum Beispiel Reliabilität und Validität. Die Reliabilität im Sinne der Interrater-Reliabilität wird dadurch optimiert, dass die Kodierung oder Bearbeitung minimale Interpretationsleistung vom Beobachter erfordert. Üblicherweise hat ein solches System einen dimensional Aufbau, das heißt, entsprechend der theoretischen Basis sind die Items oder Beob-

achtungseinheiten Teilbereichen des Verhaltens zuzuordnen, die idealerweise überschneidungsfrei und vollständig sein sollten, um abhängige Daten zu vermeiden. Die Validität wird in der Regel als Kriteriumsvalidität bestimmt, also über den Zusammenhang mit Führungserfolg.

Es gibt inzwischen sehr differenzierte und empirisch mehrfach geprüfte Taxonomien wie zum Beispiel den LBDQ (Leader Behavior Description Questionnaire) von Stogdill (1963) oder das LOS (Leader Observation System) von Luthans und Lockwood (1984). Dennoch werden nach wie vor sehr einfache, zweidimensionale Kategoriensysteme wie die Unterscheidung von aufgaben- und mitarbeiterorientiertem Verhalten bevorzugt, obwohl ihre Nachteile bekannt sind, die vor allem in der unzulänglichen empirischen Basis und dem geringen Erkenntnisgewinn durch zu starke Verallgemeinerung der Beobachtungsdaten liegen (Yukl, 1994).

3.2.6 Integrative Ansätze

Angesichts der großen Menge verschiedener Führungsmodelle, Befunde und Zugänge wurden seit Ende der achtziger Jahre integrative Ansätze gefordert, welche die Facetten der vorhandenen Ansätze abdecken sollen. Dies geschah auf mehreren Wegen. Entweder erfolgte eine rein literaturgestützte Zusammenfassung, oder die gesammelte und aufbereitete Befundlage wurde induktiv durch Expertenurteile verdichtet.

Ramme (1990) nimmt eine sehr ausführliche Literaturanalyse in Form eines strukturierten, narrativen Reviews vor, die den Ausgangspunkt für eigene Erhebungen zu den Aktivitäten von Führungskräften bildet. Dabei unterscheidet sie zunächst die Literatur zum Thema Führung nach vier Aspekten. Funktion bezieht sich auf die Beschreibung der Arbeit von Führungskräften im Sinne von Management-Funktionen. Aktivitätsorientierte Studien zielen mehr auf sichtbare Tätigkeiten wie Telefonieren, Schreiben oder Teamsitzungen ab. Mit dem Rollenkonzept werden Anforderungsprofile bezüglich verschiedener Aktivitäten umschrieben, bei der Aufgabe sollen Aktivitäten bestimmten Funktionen zugeordnet werden.

Ramme (1990) listet die Inhalte dieser Arbeiten in Form von Führungsdimensionen oder -aktivitäten tabellarisch auf, ohne die Befunde zu aggregieren. Für sie steht die Diskussion der Methoden dieser Forschungsrichtungen im Vordergrund, aus der sie unter methodischen und ökonomischen Gesichtspunkten eine optimale Kombination von Verfahren und Zugängen für ihre eigenen Erhebungen ableitet. Dabei interessiert sie, welche Beziehungen zwischen Funktions- und Fachaufgaben auf der einen und Aktivitäten auf der anderen Seite bestehen. Zu einer integrativen Verdichtung sowie zur Theoriebildung trägt diese Arbeit jedoch wenig bei, da sie nur die Hauptdimensionen

der referierten Arbeiten darstellt, nicht deren Bedeutung, die ja von Arbeit zu Arbeit selbst bei gleicher Bezeichnung einer Führungsdimension variieren können.

Schirmer (1992) referiert im ersten Teil seiner Arbeit ebenfalls diese Konzepte und empirischen Befunde, wobei er seine zusammenfassende Darstellung im Detail etwas anders als Ramme (1990) strukturiert, aber ebenfalls nicht systematisch aggregiert. Die Befundlage betrachtet er als widersprüchlich und unbefriedigend. Für ihn sind die Unstimmigkeiten dieser Befunde nicht integrierbar, sondern nur durch einen ergänzenden Zugang aufzulösen, den er in subjektiven Theorien von Führungskräften über ihren Arbeitsalltag sieht, die in funktionalen und Aktivitätsstudien nicht erfasst werden. Genauso wie bei Ramme lässt sich aus dieser Arbeit nicht auf die Anforderungen schließen, die eine Führungs- oder Managementposition an ihren Inhaber stellt.

Dagegen ist Yukls (1989) Taxonomie stärker auf eine systematische Integration ausgerichtet und erhebt den Anspruch, theoriegeleitet vorzugehen. Das geschieht, indem die Führungsdimensionen verschiedener Ansätze nach ihren Bedeutungsinhalten zusammengefasst werden, wobei sie auf der Basis von Expertenratings und Faktorenanalysen letztlich zu 14 *Managerial Practices* (Yukl, 1994) verdichtet werden. Das Vorgehen kann höchstens deshalb als theoriegeleitet angesehen werden, weil das ausgewertete Material, das selbst induktiv zusammengefasst wird, in verschiedenen Beiträgen zur Führungsforschung besteht.

Ähnlich ist das Vorgehen bei anderen Arbeiten der neunziger Jahre, die eine integrative Taxonomie von Führungskompetenzen anstreben. So sollen bei Borman und Brush (1993) ebenso wie bei Tett, Guterman, Bleier und Murphy (2000) Experten sehr viele, sehr spezifische Eigenschaftsanforderungen an Führungskräfte nach ihrer Ähnlichkeit sortieren. Bei den Experten handelt es sich um Personen, die sich wissenschaftlich oder praktisch mit Führung beschäftigen. Auch hier wurden die Anforderungen, nach denen die Befunde sortiert wurden, ausgewählten Arbeiten entnommen, und anschließend faktorenanalytisch verdichtet.

Allerdings bleibt bei diesem Vorgehen die Frage offen, inwieweit die Ergebnisse, die den Anspruch haben, spezifisch und vollständig zu sein, in ihrer Struktur nicht subjektive Führungstheorien (vergleiche Abschnitt 3.4.5) der jeweiligen Experten reflektieren. Wenigstens sind die Begrifflichkeiten vorgegeben, die auf der Grundlage vieler Arbeiten annähernd vollständig sein sollten, so dass anders als bei vielen explorativen Expertenbefragungen über Führungseigenschaften nicht die Gefahr besteht, dass wesentliche Aspekte übersehen oder Begriffe ungenau verwendet werden.

Um die Vielfalt der Befunde systematisch und quantitativ zu integrieren,

werden seit einigen Jahren in der Führungsforschung Metaanalysen durchgeführt. Dabei handelt es sich um einen Sammelbegriff für statistische Methoden innerhalb der kumulativen Analyseverfahren, die das Ziel haben, eine größere Menge bereits vorliegender Forschungsbefunde zu einem Thema zu integrieren. Die Metaanalyse kann lediglich als ökonomische Form der zusammengefassten Präsentation vorliegender Untersuchungsbefunde gesehen werden, oder zur weitergehenden Interpretation und Gesamtbeurteilung von Untersuchungen eingesetzt werden.

Ebenso wie in anderen Bereichen werden Metaanalysen auch in der Führungsforschung zunehmend kritisiert. Bereits 1994 stellt Klenke in einem Übersichtsartikel fest, dass sich Metaanalysen in der Führungsforschung typischerweise mit einer von drei Fragestellungen beschäftigen. Entweder behandeln sie Führungstheorien, Führung in Kleingruppen oder Geschlechtsunterschiede. Die Metaanalysen erscheinen nach einer systematischen Durchsicht bezüglich ihrer Methoden und Ergebnisse als genauso inkonsistent wie die Einzelstudien, und können die widersprüchliche Befundlage nicht erklären. Generell spielen Metaanalysen in der Führungsforschung keine so große Rolle wie in der Eignungsdiagnostik.

3.3 Klassische Ansätze

Auch wenn dieser Arbeit ein recht weit gefasstes Verständnis von Führung beziehungsweise Management zugrunde liegt, sollen hier die bekanntesten Ansätze zur Beschreibung von Führungsverhalten und Erklärung von Führungserfolg charakterisiert werden, die aus ihrer jeweiligen Perspektive meist nur einzelne Aspekte des Phänomens enthalten. Dieser begrenzte Blickwinkel und die damit verbundene unbefriedigende empirische Basis machen diese Modelle angreifbar und führen dazu, dass inzwischen sehr viele davon existieren, aber nicht in Einklang zu bringen sind. Gleichzeitig fehlen Ansätze, die diese Modelle und die vielen, teilweise widersprüchlichen Befunde integrieren könnten.

In der Literatur finden sich verschiedene Möglichkeiten, Führungstheorien nach bestimmten Aspekten zu gliedern. Solche Einteilungsprinzipien sind nicht nur prinzipiell arbiträr, sondern die jeweilige Zuordnung einzelner Ansätze in eine bestimmte Kategorie lässt sich gelegentlich in Frage stellen. Das gilt genauso für das hier verwendete Einteilungsprinzip, das auf verschiedene Zuordnungen aus der Führungsliteratur zurückgreift. Dabei werden diese verschiedenen Ansätze nach ihrer Orientierung an verhaltenswissenschaftlichen Forschungsparadigmen geordnet.

3.3.1 **Traittheorien**

Bei traittheoretischen Ansätzen stehen einzelne Persönlichkeitseigenschaften einer Führungskraft im Mittelpunkt. Demzufolge zeichnen sich Führungskräfte gegenüber anderen Personen durch einzelne oder mehrere besonders ausgeprägte Eigenschaften aus, auf die auch ihr Erfolg zurückzuführen ist. Die ersten Ansätze in den dreißiger und vierziger Jahren waren widersprüchlich und konnten empirisch kaum gestützt werden. Insbesondere konnte anders als erwartet kein Bündel bestimmter Fähigkeiten ausgemacht werden.

Diese ersten Resultate spiegeln eher naive Persönlichkeitstheorien oder Erwartungshaltungen wider (Yukl & Van Fleet, 1992). Das ist möglicherweise auch auf die Methoden zurückzuführen, die vor allem in Vergleichen von Führungskräften mit anderen Personen hinsichtlich der Ausprägung von Persönlichkeits- und Leistungstestscores bestanden. Etwas stabilere und konsistentere Befunde liefern die Studien ab den fünfziger Jahren, die gezielt Verhaltensdaten nutzten, um Zusammenhänge mit Führungserfolg zu untersuchen.

Demnach scheinen einige Persönlichkeitseigenschaften in einer optimalen Ausprägung relevant zu sein. So haben effektive und erfolgreiche Führungskräfte auf der motivationalen Seite in großen, hierarchisch gegliederten Organisationen idealerweise ein hohes Bedürfnis nach Einflussnahme, hohe emotionale Reife, eine höhere Leistungsorientierung und ein eher schwaches Affiliationsbedürfnis, was McClelland und Boyatzis (1982) unter dem Begriff der sozialisierten Machtorientierung zusammenfassen. Außerdem verfügen sie nach weiteren Befunden über Selbstvertrauen, persönliche Integrität, Energie und Stressresistenz (Bass, 1990; Kirkpatrick & Locke, 1991).

Neben den Persönlichkeitseigenschaften wurden auch führungsrelevante Fertigkeiten identifiziert. Beispielsweise unterscheidet eine recht einfache Taxonomie von Katz (1955) drei wesentliche Fertigkeitskategorien für Führungskräfte. Diese Taxonomie hat den Anspruch, bisherige Forschungsergebnisse zu integrieren. Demnach gibt es technische Fertigkeiten, welche die Kenntnis der Produkte und Dienstleistungen des Unternehmens enthalten sowie der Arbeitsabläufe und Ausstattung, der Märkte, Kunden und Wettbewerber.

Konzeptuelle Fertigkeiten ermöglichen, komplexe Ereignisse, Entwicklungen und Chancen zu analysieren, Probleme zu lösen und abstrakt zu denken. Interpersonelle Fertigkeiten bedeuten Einfühlungsvermögen in Personen oder Gruppen, Kooperationsfähigkeit und Kommunikation. Daneben wird eine Kategorie der administrativen Fertigkeiten aufgeführt, die Managementfunktionen wie Planung, Supervision und Delegation umfasst. Wie die empirischen Befunde zeigen, deckt diese Taxonomie die wesentlichen erfolgsrelevanten Fertigkeiten ab, differenziert aber nicht nach Situationen.

3.3.2 Verhaltensorientierte Ansätze

Verhaltensorientierte Ansätze legen ihren Schwerpunkt auf das konkrete Verhalten von Führungskräften und dessen Zusammenhang mit Führungserfolg. Ihr Ziel ist hauptsächlich, typische Verhaltensmuster zu identifizieren oder Führungsverhalten nach Taxonomien zu kategorisieren. Führungsverhalten kennzeichnet die empirisch beobachtbaren aktuellen Verhaltensweisen einer Führungsperson in einer bestimmten Situation. Die typischen Erhebungsmethoden sind direkte Verhaltensbeobachtung, Interviews und Tagebücher, um möglichst repräsentativ und realistisch die Aktivitäten des Führungsalltags zu erfassen.

Demzufolge müssen Führungskräfte auf Basis unvollständiger und gleichzeitig nicht zu bewältigender Informationen Entscheidungen treffen und stehen neben ihren Mitarbeitern mit vielen anderen Personen in Kontakt, mit denen die Zusammenarbeit nicht formal geregelt ist. Die Arbeit selbst ist hektisch, unregelmäßig und fragmentiert, variiert sehr stark und hat eher reaktiven Charakter (Kaplan, 1986). Das widerspricht früheren Erwartungen, dass eine Führungskraft langfristig planen und diese Planung kontrolliert umsetzen könne.

Im Rahmen verhaltensorientierter Ansätze werden Taxonomien und Kategoriensysteme zur Klassifikation des Verhaltens entwickelt und eingesetzt. Darunter gibt es zweifaktorielle Taxonomien, die den Anspruch haben, damit zwischen erfolgreichem und ineffektivem Führungsverhalten zu unterscheiden. Die meisten davon basieren auf den Ohio-Führungsskalen, die das Führungsverhalten im Wesentlichen nach Mitarbeiterorientierung (*consideration*) und Aufgabenorientierung (*initiating structure*) einteilen. Trotz der Beliebtheit dieser Skalen liegen erstaunlich wenige Befunde zum Führungserfolg oder der Mitarbeiterleistung vor, außer dem, dass die Mitarbeiterorientierung positiv mit der Arbeitszufriedenheit der Mitarbeiter korreliert (Yukl, 1989). Neben der starken Vereinfachung ist diese Einteilung nicht trennscharf, da Führungsverhalten meist nicht allein dem Zweck der Mitarbeiterorientierung oder der Aufgabenorientierung dient.

Weitere Studien in diesem Rahmen untersuchen den Zusammenhang mit Arbeitsleistung, wonach Führungserfolg besser durch spezielle Verhaltensweisen vorhergesagt werden kann als durch globale Bewertungsskalen des Führungsverhaltens. Demzufolge können effektive Führungskräfte beispielsweise Beziehungen zwischen verschiedenen Problemfeldern herstellen und diese dadurch effizienter bewältigen. Daneben ist es förderlich, Mitarbeiter möglichst präzise und konkret zu instruieren (*clarifying*), zu belohnen und zu motivieren. Problemlöseverhalten wird ebenso als günstig angesehen wie Planung, auch wenn diese unter den Bedingungen des Führungsalltags erschwert ist.

3.3.3 Kognitive Ansätze

Als kognitive Ansätze werden hier solche Führungstheorien bezeichnet, die typische psychische Prozesse, notwendige Fertigkeiten und Fähigkeiten von Führungskräften bei ihrer Arbeit beschreiben und damit außerdem Führungserfolg erklären wollen. So stellen zum Beispiel Fleishman, Mumford, Zaccaro, Levin, Korotkin und Hein (1991) eine Übersicht der Klassifikationssystematiken dar, die Führungsverhalten erfassen. Für den Zeitraum zwischen Mitte der siebziger und Mitte der achtziger Jahre konstatieren sie als Trend, dass sich diese Modelle zunehmend auf Prozesse der Informationsverarbeitung konzentrieren, und weniger als zuvor auf funktionale und soziale Aspekte der Führung.

Diese neuen Dimensionen des Führungsverhaltens sind Kommunikation, Planung und Entscheidungsverhalten. Fleishman et al. (1991) führen diese Entwicklung auf die seit Anfang der siebziger Jahre gewachsene Bedeutung der kognitiven Psychologie zurück und auf die damit verbundene stärkere Berücksichtigung kognitiver Prozesse in sozialen Situationen wie bei Bandura (1986) und anderen grundlagenwissenschaftlichen Ansätzen zu sehen ist. Handlungstheorien ergänzen die Taxonomien von Führungsverhalten und sind Bestandteil eines erweiterten Verständnis von Führung als zielorientiertem Verhalten.

Die von Fleishman et al. (1991) entwickelte Taxonomie, die so genannten Leadership Behavior Dimensions (LBDs), ist selbst in diesen Trend einzuordnen. Sie setzt auf bisherige Beschreibungssysteme von Führungsverhalten auf und ist mit ihren Inhalten ein Produkt der damaligen Grundlagenforschung. Führungsverhalten lässt sich demnach in vier übergeordnete Dimensionen klassifizieren. *Information Search and Structuring* beinhaltet Aktivitäten zum Erwerb von Informationen, zur Organisation und Bewertung sowie Rückmeldungs- und Steuerprozesse. *Information Use in Problem Solving* umfasst die Identifikation von Handlungsbedarf, Planung, Koordination und die Weitergabe von Informationen.

Mit *Managing Personnel Resources* sind vier Bereiche angesprochen: Personalbeschaffung und -zuordnung, Personalentwicklung, Motivation sowie Einsatz und Überwachung von Mitarbeitern. Ähnliche Aktivitäten erfordert schließlich die vierte Dimension *Managing Material Resources* mit der Materialbeschaffung und -zuordnung, der Instandhaltung sowie dem Einsatz und der Überwachung der materiellen Ressourcen.

Bemerkenswert an dieser Darstellung ist, dass die beiden Dimensionen der Informationsverarbeitung Querschnittsfunktionen sowohl für den Umgang mit den Mitarbeitern oder so genannten Personalressourcen als auch mit den materiellen Ressourcen darstellen. Es werden also die gleichen kognitiven

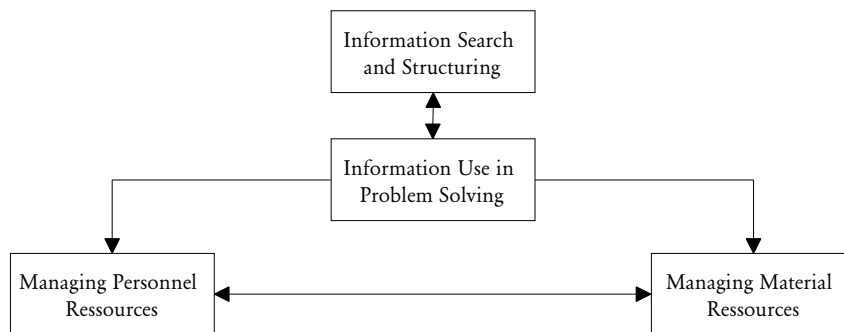


Abbildung 3.2: Beziehungen zwischen den Hauptdimensionen des LBDs (nach Fleishman et al., 1991).

Prozesse dafür angenommen, statt die sozialen und die nicht-sozialen Aktivitäten als grundsätzlich und qualitativ unterschiedlich anzusehen, wie dies in den meisten früheren Führungsmodellen der Fall ist. Die beiden übergeordneten Dimensionen der Informationsverarbeitung weisen mit ihren Inhalten eine hohe Überschneidung mit den Dimensionen des Problemlösemodells von Dörner (1986) auf.

Solche kognitiven Ansätze werden bis heute in der Führungsforschung weiterentwickelt (Marshall-Mies, Fleishman, Martin, Zaccaro, Baughman & McGee, 2000), auch die Diagnose der Problemlösefähigkeit ist nach wie vor aktuell. Die Beiträge beschäftigen sich unter anderem mit Führung als Bewältigung komplexer sozialer Probleme (Mumford, Zaccaro, Harding, Jacobs & Fleishman, 2000), mit Verfahren zur Erfassung von Problemlösefähigkeit bei Führungskräften (Zaccaro, Mumford, Connelly, Marks & Gilbert, 2000) oder mit kognitiven und meta-kognitiven Maßen zur Vorhersage von Führungspotential. Damit erscheint die Problemlöseforschung nach wie vor als ein möglicher und interessanter Zugang zur Managementdiagnostik.

3.3.4 Führungsstilforschung

Führungsstil ist im Unterschied zum Führungsverhalten ein langfristiges, situationsunabhängiges, regelmäßig wiederkehrendes Verhaltensmuster von Vorgesetzten gegenüber ihren Mitarbeitern, also die Grundausrichtung, die das Führungsverhalten kennzeichnet. Die Führungsliteratur kennt eine Reihe von Typologien zur Unterscheidung von Führungsstilen. Sie beruhen entweder auf empirischen Verhaltensbeobachtungen, die in diesem Sinne abstrahiert werden, oder auf Idealtypologien, die auf der Grundlage von Theorien oder Plausibilitätserwägungen entwickelt wurden. Diese Typologien sind

durch bestimmte Ausprägungen auf verschiedenen Dimensionen charakterisiert. In der Regel bestehen diese Dimensionen in der Aufgabenorientierung und Mitarbeiterorientierung der Vorgesetzten sowie in der Partizipationsmöglichkeit der Mitarbeiter.

Lewins Ansatz ist der älteste der Führungsstilforschung (Lewin, Lippitt & White, 1939) und beruht auf empirischen Untersuchungen mit Kindern, die von Aufsichtspersonen mit unterschiedlichem Führungsstil betreut wurden. Aus dem praktizierten Führungsstil resultieren demzufolge bestimmte Verhaltensweisen der Gruppenmitglieder. Lewin unterscheidet dabei drei wesentliche Führungsstile mit abnehmendem Grad an Einflussnahme und zunehmender Mitbestimmungsmöglichkeit.

Beim autoritären Führungsstil lenkt die Führungskraft die Aktivitäten und Ziele des Einzelnen und der Gruppe. Ein Vorgesetzter mit demokratischem Führungsstil ermuntert die Gruppenmitglieder zur Initiative und gemeinsamen Entscheidung, nimmt aber auch selbst aktiv am Gruppenprozess teil. Hat die Gruppe völlige Entscheidungsfreiheit und greift der Vorgesetzte nicht steuernd ein, spricht Lewin vom laissez-faire-Führungsstil. Generell soll ein demokratischer Führungsstil ideal sein. Trotz der empirischen Basis dieses Ansatzes konnte dies in Folgeuntersuchungen nicht bestätigt werden.

Daneben gibt es idealtypische Darstellungen von Führungsstilen, die eine, zwei oder mehr Dimensionen postulieren. Tannenbaum und Schmidt (1958) unterscheiden ein siebenfach abgestuftes Kontinuum von Verhaltensklassen nach dem Grad der Einbeziehung der Mitarbeiter. Dies reicht von autoritärem Führungsstil ohne Mitbestimmungsmöglichkeit bis zum demokratischen Führungsstil, wobei die Gruppe allein entscheidet und der Vorgesetzte nur noch koordiniert.

Blake und Mouton (1964) gehen in ihrem Managerial Grid neben der Mitarbeiterorientierung von einer weiteren Dimension aus, der Produktionsorientierung. Der Führungsstil ist durch die Kombination der Ausprägungen auf beiden Dimensionen gekennzeichnet, wobei eine hohe Ausprägung auf beiden Dimensionen optimal sein soll. Lattmann (1982) differenziert Führungsstil weiter als Profil von 15 Merkmalsausprägungen auf einer Matrix. Gemeinsame Probleme dieser Ansätze bestehen in der fehlenden empirischen Basis, der ungeklärten Frage nach der Vollständigkeit der Systematik sowie der mangelnden Klassifizierbarkeit aller denkbaren oder vorkommenden Verhaltensweisen nach diesen Kriterien.

3.3.5 Interaktionistische Ansätze

Interaktionistische Ansätze stellen eine Reaktion auf die mangelnde empirische Evidenz der Eigenschaftstheorien und der Führungsstilansätze dar. Sie

gehen davon aus, dass die Wirkung des Führungsprozesses nicht nur vom Verhalten der Führungskraft, sondern außerdem von den jeweiligen spezifischen Gegebenheiten der Situation abhängen. Zu diesen Kontextfaktoren gehören die formale Autorität der Führungsperson, die Aufgabenstruktur, Eigenschaften der Mitarbeiter und der externen Umwelt.

Dabei gibt es grundsätzlich zwei Zugangsmöglichkeiten. In der ersten Variante wird das Führungsverhalten als abhängige Variable gesehen, und die Wirkung verschiedener Situationen untersucht oder man betrachtet, wie sehr das Verhalten über verschiedene Situationen variieren kann. Die alternative und häufiger gewählte Zugangsweise, der kontingenztheoretische Ansatz, betont, wie situative Variablen die Beziehung zwischen Eigenschaften oder Verhalten des Vorgesetzten und Führungserfolg moderieren.

Diese situativen Einflussfaktoren können zum Beispiel in einem Rollenmodell dargestellt werden, das die Erwartungen von Vorgesetzten, Kollegen, Mitarbeitern und Externen enthält (Kahn, Wolfe, Quinn & Snoek, 1964). Oder sie stehen in einem Modell multipler Einflüsse (Hunt, 1991), die auf der Makro-Ebene Eigenschaften der Organisation, Abteilungsgröße, Technologie und Wechselbeziehungen der externen Umwelt sind und auf der Mikro-Ebene unmittelbare Situationsvariablen darstellen, wie zum Beispiel Aufgabenstruktur, Einstellung der Mitarbeiter und Gruppenzusammenhalt. Sinnvoll ist es auch, zu untersuchen, wie Führungskräfte bestimmte Situationen wahrnehmen und damit umgehen, und wie effektiv diese Reaktionen sind, was wieder eine Verbindung zum Eigenschaftsansatz herstellt. Dazu gibt es bisher nur wenige Untersuchungen.

Wesentlich mehr kontingenztheoretische Studien und Ansätze beschreiben, welche Aspekte der Situation die Beziehung zwischen Führungsverhalten oder Eigenschaften und dem Ergebnis bestimmen. Die Path goal theory (Evans, 1970) nimmt an, dass Führungskräfte ihre Mitarbeiter zu höheren Leistungen motivieren können, wenn ihnen vermittelt wird, dass sie durch ihre Anstrengungen wertvolle Resultate erzielen.

Nach der Führungstheorie von Hersey und Blanchard (1969) hängt die optimale Kombination von Aufgaben- und Mitarbeiterorientierung von der Reife der Mitarbeiter ab, wobei der Begriff der Reife unklar bleibt. Kerr und Jemier (1978) gehen davon aus, dass bestimmte situative Faktoren den Einfluss von formalen Führungskräften reduzieren, indem sie diese überflüssig machen oder die Wirkung ihrer Führungsaktivitäten neutralisieren. Vroom und Yetton (1973) unterscheiden fünf unterschiedlich partizipative Entscheidungsstrategien, die in Abhängigkeit von sieben situativen Voraussetzungen und weiteren Zielvorgaben gewählt werden.

In seiner Kontingenztheorie nimmt Fiedler (1967) an, dass in Abhängigkeit von der sozioemotionalen Beziehung zu den Mitarbeitern, der Aufga-

benstruktur und der Sanktionsmacht ein aufgaben- oder personenorientierter Führungsstil zu bevorzugen ist. Unter welchen Umständen in Form von Stress, Aufgabenstruktur und Führungsstil die kognitive Leistungsfähigkeit und Fachkompetenz die Gruppenleistung beeinflussen, präzisiert die Theorie kognitiver Ressourcen von Fiedler und Garcia (1987).

Das Modell multipler Beziehungen (Yukl, 1994) führt die Arbeitsleistung einer Abteilung auf sechs Variablen der Aufgaben, der Gruppe und auf externe Faktoren zurück, die wiederum direkt durch situative Faktoren beeinflusst oder dadurch in ihrer Bedeutung gewichtet werden. Vorgesetzte können auf beide Faktorengruppen Einfluss nehmen. Ähnliches enthält die Theorie zur Vorgesetzten-Umwelt-Mitarbeiter-Interaktion (Wofford, 1982), gibt aber genauere Empfehlungen für Maßnahmen des Vorgesetzten.

Weil die meisten interaktionistischen Ansätze sehr komplex sind, ist es schwierig, spezifische Hypothesen abzuleiten und empirisch zu testen, da die theoretischen Annahmen global formuliert und abhängige, intervenierende und unabhängige Variablen schwer zu operationalisieren sind, beziehungsweise weil die Operationalisierungen und damit verbundenen Ergebnisse notwendigerweise kaum zu interpretieren sind. Auch wird meist nicht spezifiziert, wie die intervenierenden Variablen genau wirken sollen. Demzufolge sind die wenigen empirischen Befunde, die es dazu gibt, uneinheitlich und führen nicht weiter, wenn es um die Identifikation verhaltens- oder leistungsdeterminierender Faktoren geht.

3.3.6 Rollenansätze

Obwohl die rein deskriptiven Tätigkeitsbeobachtungen bis heute vorherrschen, wurden sie zunehmend wegen ihrer mangelnden Fruchtbarkeit für die Theoriebildung sowie wegen ihrer fehlenden konzeptuellen Fundierung kritisiert. Mintzberg (1973) versucht in einem integrativen Ansatz sowohl die Schwächen des funktionalen Ansatz wie des empirischen Vorgehens zu vermeiden und gleichzeitig die Stärken beider zu verbinden. Mit dem Rollenkonzept, einer personenunabhängigen Verhaltenserwartung, können die Aktivitäten gleichzeitig beobachtet und in einen sozialen und organisationalen Kontext eingebettet werden. Zu diesem Zweck verdichtet Mintzberg auf der Merkmalsebene die Daten über die Aktivitäten zu zehn idealtypischen Rollen, mit denen die Arbeit von Führungskräften erschöpfend beschrieben werden können. Auf der Objektebene werden wiederum acht Jobtypen vermutet, die jeweils durch Schlüsselrollen charakterisiert werden.

Trotz der eigenen Ansprüche ist die empirische Basis für Mintzbergs Konzept sehr schmal, was sich auch daran zeigt, dass die Replikationen, zu denen es anregte, sehr widersprüchlich ausfielen (Pavett & Lau, 1983). Zu-

dem unternahm in den Folgestudien kaum jemand eine empirisch gestützte Weiterentwicklung der Jobtypen. Der Wert lag eher in der Anregung neuer, weiterführender Untersuchungsschwerpunkte durch das Rollenmodell. Dieses wurde als konzeptueller Rahmen zur Entwicklung von Beobachtungsinstrumenten zur Erfassung von Arbeitsinhalten genutzt. Die daraus resultierenden Daten wurden wiederum nach Arbeitsinhalten faktorisiert und als weiteres Novum auf ihren Vorhersagewert für Managereffektivität anhand ökonomischer Kriterien für verschiedene Branchen geprüft (Morse & Wagner, 1978).

Neuberger (1990) weist darauf hin, dass der Rollenansatz kein einheitliches Theoriegebäude ist, sondern sich unter verschiedenen Aspekten betrachten lässt, die jedoch alle mit theoretischen und praktischen Schwächen behaftet sind. Den Ansatz von Mintzberg (1973) bezeichnet er als „Rückfall in elitäre Great Man-Phantasien“ (Seite 100), weil sich diese Rollen allein auf den Vorgesetzten konzentrieren. Gleichzeitig konstatiert er einen internen Widerspruch zwischen der gleichzeitig postulierten Autonomie von Führungskräften und der Fremdbestimmung durch Rollenerwartungen. Letzten Endes kann man, auch angesichts der Befundlage, nach Rosenstiel (2001) diesen Ansatz als spekulativ bezeichnen.

3.4 Neuere Ansätze

Neben diesen Perspektiven gibt es neuere Entwicklungen in der Führungsforschung, die ebenfalls für Praxis und Theoriebildung relevant sind, ohne dass sie sich in die bisher dargestellten Ansätze einordnen lassen, die verhaltenswissenschaftlichen Forschungsrichtungen entsprechen. Vor dem Hintergrund der Erkenntnis, dass die beschriebenen Ansätze nicht ausreichen, den Führungsalltag abzubilden, versuchen neuere organisationstheoretische Konzeptionen stärker, der Komplexität und den sozialen Prozessen des Führungsalltags gerecht zu werden, die eher als Theorien des Geführtwerdens sens. Neuberger (1990) bezeichnet werden könnten. In eine ähnliche Richtung gehen Wunderers (1997) Überlegungen zu einer Führung „von unten“, die er von einer Führung von oben im gängigen Sinne sowie von der lateralen Kooperation abgrenzt. Neben den Einflussgrößen auf diesen Bedeutungswandel beschreibt und prüft er erfolgreiche Strategien.

3.4.1 Systemische Führungstheorie

Organisationen und deren Umwelt sind offensichtlich durch viele und vernetzte Zusammenhangsbeziehungen geprägt, die zudem dynamisch und rückbezüglich sein können. Das hat dazu geführt, dass Organisationen und organi-

sationale Prozesse häufig in Systemmodellen dargestellt werden. Neuberger (1990) weist darauf hin, dass die Bezeichnung *Systemische Führungstheorien* einen Widerspruch in sich birgt, da Führung eine gezielte Steuerung von außen darstellt, während Systeme durch ungeleitete Entwicklungen von innen charakterisiert sind. Organisationale Systemtheorien greifen auf technische, naturwissenschaftliche oder soziologische Systemmodelle zurück, die einige grundlegenden Gemeinsamkeiten aufweisen, aber sich in ihrer Terminologie und einigen Annahmen unterscheiden.

Technische Systemtheorien (Kybernetik) stammen aus den Ingenieurwissenschaften und befassen sich mit der Steuerung mechanischer Systeme durch Regelprozesse, die dazu führen, dass Ausgangsvariablen, zum Beispiel bei einem Thermostaten die Zimmertemperatur auf konstantem Niveau bleibt. Dazu gehören formale Modelle wie das der finiten Automaten (Buchner, 1999). Biologische Systemansätze versuchen, das Zusammenspiel verschiedener Funktionen lebender Organismen abzubilden, die aus Teilsystemen bestehen. Besondere Aufmerksamkeit gilt der Reaktion auf Veränderungen der Umwelt, und den Prozessen, die zur Stabilisierung des Systemzustandes führen (Homöostase) oder zu einer Veränderung (Heterostase). Pathologische Entwicklungen lassen sich damit ebenfalls modellieren.

In eine ähnliche Richtung gehen evolutionäre Organisationstheorien, die Organisationen und ihre Prozesse analog zu biologischen Systemen abbilden und insbesondere Prinzipien der Entwicklung und Veränderung aus der Evolutionstheorie übertragen. Die Anwendung von Evolutionsprinzipien ist jedoch die wesentliche Gemeinsamkeit. Darüber hinaus unterscheiden sich diese Ansätze erheblich, zum Beispiel bezüglich der Einheiten, für die Evolution stattfindet, und bezüglich der Einheiten, die als Umwelt definiert werden (Kieser, 1997). Wichtige Prozesse sind dabei zum Beispiel Mutation, Adaption oder Selektion, die per Definition ungeleitet stattfinden und zu Veränderungen von Systemen durch Umweltbedingungen führen.

Sozialwissenschaftliche Systemtheorien übertragen diese Modelle auf soziale Prozesse und Systeme, um deren Komplexität gerecht zu werden, wobei sie systemtheoretische Begriffe teilweise neu definieren und zudem eigene Bezeichnungen einführen. Luhmann (1984) ist der bekannteste Vertreter dieses Ansatzes. In seinem Modell, das immer wieder überarbeitet und weiterentwickelt wurde, definiert er nicht nur Eigenschaften sozialer Systeme, sondern auch die Konsequenzen für das Verhalten darin. Dabei ist Komplexität durch die Anzahl von Elementen und Beziehungen sowie der Unterschiedlichkeit von Beziehungen definiert und kann auf das System sowie seine Umwelt angewandt werden. Insbesondere im Umgang mit der Umwelt muss die Komplexität durch Selektion der möglichen Beziehungen reduziert werden.

Kontingenz bezeichnet einen Zustand mit den darauf bezogenen Ände-

rungsmöglichkeiten. Soziale Interaktionen sind durch doppelte Kontingenz charakterisiert, das heißt dass die Teilnehmer ihren Ausgangszustand und den ihres Gegenübers vor Augen haben sowie die damit verbundenen eigenen Handlungsmöglichkeiten und die des Gegenübers. Autopoiese beschreibt den Sachverhalt, dass sich Systeme aus sich selbst erzeugen und aufrechterhalten können, und im sozialen System Systemstrukturen definiert und verwendet werden (Luhmann, 1988). Aus diesen Modellvorstellungen resultierten Verhaltensempfehlungen zum Umgang mit der Komplexität, wie sie zum Beispiel Malik (1986) formuliert, jedoch auch die Einschätzung, dass soziale Systeme zu komplex seien, um noch gesteuert zu werden.

In diesem Kontext vertritt Türk (1981) mit seinem erweiterten systemtheoretischen Ansatz aus der Soziologie die These, dass Personalführung nur ein Residualfaktor sei, da soziale Handlungen bereits durch Prozesse aus dem Organisationssystem, dem Interaktionssystem und dem Personensystem gesteuert seien. Zudem sei Personalführung nur ein funktionales Äquivalent von sozialer Kontrolle, aber nicht notwendigerweise die einzige Ursache.

Systemtheoretische Modelle haben den Vorteil, eine neue und differenziertere Perspektive auf Führung in Organisationen zu eröffnen, die den vielfältigen und vernetzten Vorgängen des Alltags eher gerecht werden als die klassischen Managementtheorien, da sie den Focus von der Führungskraft auf das Gesamtsystem erweitern. Ihre Stärke liegt eindeutig in der Erklärung von Verhalten als ein Ziel psychologischer Forschung. Weniger gut können sie Verhalten und Verhaltensergebnisse vorhersagen. Sie ermöglichen eine aufschlussreiche Analyse im Nachhinein, wie sie etwa Neuberger (2000a) zur Erklärung der fehlgeschlagenen Fusion zwischen Deutscher Bank und Dresdener Bank vornimmt.

Gleichzeitig ist es schwierig, im Rahmen systemtheoretischer Ansätze prüf- und falsifizierbare Hypothesen eindeutig abzuleiten, da sich jede beliebige Entwicklung modellieren lässt, was Systemtheorien gegen Falsifikation immunisiert. Ein weiteres Problem ist die bisher ungeklärte Definition von Komplexität (Casti, 1979) in diesen und allen anderen Ansätzen, die mit dem Konzept arbeiten. Aus beidem resultieren auch widersprüchliche Handlungsempfehlungen, da es vorher nicht eindeutig zu entscheiden ist, ob es zum Beispiel günstiger ist, die Stabilität oder die Adaptivität eines Systems zu optimieren, und im Nachhinein nicht festgestellt werden kann, ob die Alternative nicht doch besser gewesen wäre. Damit lenkt die Systemperspektive den Blick auf personale Voraussetzungen zum Umgang mit Komplexität und Unsicherheit, die in den Eigenschaftskonstrukten Problemlösefähigkeit und Ambiguitätstoleranz enthalten sind.

3.4.2 Mikropolitik

Die Aktivitätsforschung zeigte nicht nur die Grenzen funktionaler Managementmodelle auf, sondern lieferte Hinweise, dass noch andere Verhaltensweisen zum Alltag von Führungskräften gehören, die dort nicht aufgeführt sind. Vor allem Aktivitäten, welche den Aufbau und Einsatz von persönlichen Beziehungen betreffen, werden als relevant für organisationale Prozesse betrachtet. Weiterhin werden sie nicht nur von Führungskräften genutzt, und Führungskräfte sind nicht nur Handelnde, sondern ebenso Ziel und Ressource. Diese sind unter dem Begriff Mikropolitik gefasst, da sie im Unterschied zur staatlichen und zur Unternehmenspolitik die organisationale Innenpolitik prägen (Küpper & Ortmann, 1988).

Dabei reicht das Begriffsverständnis hier von einer Ausnutzung anderer durch einen bestimmten Persönlichkeitstyp über Manipulation und mikropolitische, informelle Verhaltensweisen jenseits der offiziellen Politik bis hin zu einem organisationstheoretischen Konzept. Entsprechend wird Mikropolitik aus verschiedenen Perspektiven behandelt, es gibt individualpsychologische Ansätze, sie können als Machttheorien oder im Rahmen strategischer Organisationsanalysen betrachtet werden. Mikropolitik ist damit eine strategische Nutzung interessierter, mehr oder minder mächtiger Akteure von organisationalen Strukturen, wodurch sie diese Strukturen beabsichtigt oder unbeabsichtigt verfestigen, untergraben oder verändern können.

Weiterhin wird ein aspektuales Verständnis der Mikropolitik von einem konzeptualen Verständnis unterschieden (Küpper & Felsch, 2000). Im ersten Fall wird Mikropolitik als ein Aspekt neben anderen Methoden der Handlungssteuerung aufgefasst. Nicht jedes Handeln ist nach dieser Sichtweise mikropolitisch, sondern nur, wenn auf illegitime Weise politische Ziele verfolgt werden. Demzufolge kommt Mikropolitik in Unternehmen unterschiedlich oft vor, daneben gibt es einen unpolitischen Raum. Die konzeptuale Sichtweise betrachtet dagegen alle Interaktionen als interessengeleitet und politisch. Der Fokus ist nicht auf einen besonders machtgierigen Menschentypen gerichtet, sondern vielmehr sei politisches Handeln ein normales Phänomen. Konflikte entstehen durch die normalen Ziele der Identitätsbehauptung und Interessendurchsetzung. Mikropolitik ist hier nicht mit einem negativen Akzent versehen.

Eng damit verbunden ist also Machtausübung oder Einflussnahme (Blickle & Hepperle, 1999). Die Führungsforschung beschäftigt sich damit, aus welchen Quellen diese Macht stammt, wie man sie gewinnt und verliert und wie sich Vorgesetzte und Mitarbeiter gegenseitig beeinflussen. Weitere Felder sind die Auswirkungen von Macht auf Führungserfolg und Taktiken zur Einflussnahme. Die bekannte Taxonomie von French und Raven (1959)

unterscheidet fünf Arten von Macht: Legitimationsmacht, Sanktionsmacht, Ausführungsmacht, Expertenmacht und Referenzmacht. Später wurde sie um zwei weitere Machtquellen ergänzt, um die Informationsmacht und den Einfluss auf wichtige Personen (Yukl & Falbe, 1991). Aus einer anderen Perspektive können die Machtquellen danach eingeteilt werden, ob sie auf Eigenschaften einer Person (persönliche Macht) oder der Situation (Positionsmacht) beruhen (Bass, 1990).

Als Inhalte von Mikropolitik werden auch häufig konkrete Taktiken beispielhaft oder möglichst erschöpfend aufgeführt. Neuberger (1995) fasst diese in sieben Bereiche zusammen, die er aus anderen Untersuchungen, populärwissenschaftlichen Publikationen, Seminaren und Fallstudien extrahiert. Beispiel dafür sind Informationskontrolle, Beziehungspflege oder Selbstdarstellung. Ungeklärt bleibt, wie groß der Anteil solcher Taktiken im Alltag ist, wie sie genau funktionieren und welche Folgen sie haben. Das liegt auch daran, dass Mikropolitik nicht nur schwer zu definieren, sondern auch empirisch schwer zu erfassen ist, da sie laut Definition nur dann wirksam werden kann, wenn sie von den Beteiligten nicht explizit gemacht wird.

Deshalb kann nur darüber spekuliert werden, in welchem Umfang der Erfolg von Führungskräften davon abhängt, und ob dies eher der persönlichen Karriere oder den Unternehmenszielen dient. Die aspektuale Definition von Führungspolitik, die üblicherweise herangezogen wird, um die Relevanz von Mikropolitik hervorzuheben, impliziert, dass alles Handeln politisch ist. Damit gibt es nichts mehr, was nicht Mikropolitik ist. Wissenschaftstheoretisch gesehen, hat dies einen Informationsgehalt von Null, oder, wie es Gebert (1995) formuliert, eine solche Auffassung „erweist sich damit als Definitionsartefakt“ (Seite 69). Zudem leidet die Effizienz von Betrieben schon allein unter dem Bewusstsein über die Unvermeidlichkeit von Mikropolitik, die etwa Neuberger (1995) postuliert. Ob man dann noch davon ausgehen kann, dass sie nützlich ist, wie Neuberger (1995) annimmt, ist fraglich.

3.4.3 Symbolische Führung

Das Modell der symbolischen Führung (Neuberger, 1990) enthält zwei Aspekte. Zum einen drückt sich in jeder Organisation und ihren Elementen Führung aus und nimmt damit auf den einzelnen Einfluss. Zum anderen wirken Vorgesetzte mit ihren Handlungen vermittelnd oder symbolisierend, indem die Handlungen von anderen gedeutet werden und geregeltes Anschlusshandeln auslösen. Damit haben sie keine aktive Rolle, sondern fungieren ebenfalls symbolisch. Noch stärker als die mikropolitischen Ansätze stellt diese Sichtweise die bewusst steuernde Funktion von Führungskräften in Frage. Damit besteht symbolisches Management darin, einerseits aus Symbolen Fak-

ten zu machen und andererseits Fakten als Symbole anzusehen.

Bei der symbolisierten Führung (Lasser, 1987) erfolgt die Steuerung auf verschiedenen Wegen, schriftlich, über Gewohnheiten des sozialen Umgangs und über Gegenstände. Sie wirken personenunabhängig, weil sie als Fakten existieren und sich an ihnen die Handlungen orientieren. Bei symbolisierter Führung geht es vor allem um das Neuinterpretieren alter Fakten und die Schaffung neuer Fakten und sowie deren Deutung. Dementsprechend müssen hermeneutische und phänomenologische Verfahren symbolischer Führung empirisch erfassen, weil nicht nur die wichtigen Elemente der symbolischen Führung beobachtet werden sollen, sondern auch deren Deutung. Wichtig ist hierbei, dass neben den Führungskräften die Perspektive der Mitarbeiter einbezogen wird. Das Konzept hat viele Überschneidungspunkte mit Organisationskultur.

Kritik an der symbolischen Führung betrifft die praktische Bedeutung und Anwendbarkeit dieses Ansatzes. So meint auch Rosenstiel (1995), Untersuchungen zur symbolischen Führung wollen zwar den Horizont erweitern, Hilfe für Führende geben sie jedoch nicht, auch als Sozialtechnologie sei der Ansatz nicht brauchbar. Zudem erscheine die empirische Überprüfung der Theorie schwierig, da diese vorwiegend theoretisch argumentiert. Neuberger (1990) widerspricht dieser Kritik, da die Theorie durchaus Heuristiken biete, mit denen der Führungsprozess gesteuert werden kann. Dennoch bietet sie dem Vorgesetzten letzten Endes nicht viel mehr an, als möglichst genau hinzusehen und den verborgenen Sinn von Handlungen und Fakten beim Führungshandeln zu berücksichtigen.

3.4.4 Charismatische und transformationale Führung

Obwohl bereits Weber (1956) charismatische Führung konzeptualisierte, gewann sie erst in den achtziger Jahren an Bedeutung, als Unternehmen unter großen Wettbewerbsdruck gerieten und auch im Führungsbereich nach neuen Modellen suchten. Zwischen charismatischer und transformationaler Führung besteht ein großer Überschneidungsbereich. Transformationale Führung richtet sich auf die Beeinflussung der Organisationskultur und gewinnt die Unterstützung der Organisationsmitglieder bei größeren Veränderungen der Organisationsziele und Strategien. Charismatische Führung bezieht sich dagegen mehr auf das Verhältnis zwischen Führungskräften und Mitarbeitern, die ihren Vorgesetzten als herausragende Persönlichkeit wahrnehmen und verehren. Damit ist charismatische Führung durch die Eigenschaften des Vorgesetzten charakterisiert, wobei Führung aus einer breiteren Perspektive gesehen wird als durch die Traittheorien der Führung.

Diese neueren Führungsmodelle fokussieren besonders auf Prozesse, die

mit sozialen Aspekten von Führung zusammenhängen, insbesondere der dyadischen Beziehung zwischen Führungskraft und Mitarbeiter. Sie benennen Variablen, die zu der besonderen Wahrnehmung des Vorgesetzten durch den Mitarbeiter führen, was eine meist günstige Verhaltensänderung zur Folge hat. Die empirische Basis ist nach wissenschaftlichen Kriterien eher dünn und bestand anfangs nur in anekdotischen Einzelfallanalysen prominenter Persönlichkeiten. Inzwischen gibt es zunehmend Aktivitäten, die Dimensionen dieser Modelle zu diagnostizieren und auf eine breitere Datenbasis zu stellen (zum Beispiel Carless, 1998). Sie werden in der Fachliteratur sehr kontrovers diskutiert.

Zur transformationalen und charismatischen Führung gibt es verschiedene Theorien. Der Ansatz von House und Howell (1992) befasst sich mit dem Verhalten von charismatischen Führungskräften, den Eigenschaften, in denen sie sich von anderen Personen unterscheiden, und den Bedingungen, unter denen sie am besten zur Geltung kommen. Die Verhaltensweisen bestehen in impression management, Kommunikation von Visionen, hohen Erwartungen und Vertrauen an die Mitarbeiter. Als besondere Eigenschaften werden ein starkes Machtbedürfnis, großes Selbstvertrauen und starke Überzeugungen genannt.

Conger und Kanungo (1987) gehen davon aus, dass das Phänomen der charismatischen Führung durch Attributionsprozesse der Mitarbeiter entsteht, welche wiederum durch bestimmte Verhaltensweisen und Eigenschaften des Vorgesetzten begünstigt werden. Solche Verhaltensweisen bestehen darin, eine besonders ansprechende und neuartigen Vision begeistert zu vertreten, dafür persönliche Risiken einzugehen, Ressourcen für die Umsetzung zu investieren sowie unkonventionelle Durchsetzungsstrategien einzusetzen. Charismatische Vorgesetzte verfügen auf der Eigenschaftsseite über Selbstvertrauen, Techniken des Impression managements, kognitive Fähigkeiten zur adäquaten Situationseinschätzung und soziales Einfühlungsvermögen.

Nach Burns (1978) ist transformationale Führung ein Prozess, der auf Mikroebene die Einflussnahme von Individuen beinhaltet und auf der Makroebene die Veränderung von sozialen Systemen und institutionelle Reformen mobilisiert. Das geschieht dadurch, dass an die höheren Werte anderer Personen appelliert wird, wozu prinzipiell jeder in jeder Situation fähig sein sollte. Auf diesem Ansatz baut Bass (1985) auf, der außerdem zwischen charismatischer, transformationaler und transaktionaler Führung unterscheidet.

Transformationale Führung besteht in der Wirkung von Führungskräften auf ihre Mitarbeiter, welche die vom Vorgesetzten vermittelten Werte verinnerlichen, sich dadurch verändern und stärker engagieren. Charisma dagegen ist ein Prozess, der bei den Mitarbeitern starke Emotionen und Identifikation mit dem Vorgesetzten bewirkt, was eine notwendige, aber keine hinreichen-

de Bedingung für transaktionale Führung ist, die außerdem intellektuelle Anregung und persönliche Zuwendung erfordert. Transaktionale Führung fokussiert mehr auf die dyadische Beziehung zwischen Vorgesetztem und Mitarbeiter. Sie beinhaltet zusätzlich weitere Aspekte der Einflussnahme, wie etwa die Offenlegung der Leistungen, die nötig sind, um Anerkennung zu gewinnen.

Yukl (1999) diskutiert in einem Übersichtsartikel kritisch die inhaltlichen Schwächen, die Befundlage sowie die methodische Grundlage der transformationalen und der charismatischen Führungstheorien. Viele Kritikpunkte könnten auch auf die meisten anderen Führungstheorien angewendet werden, wie er selbst schreibt, sollen jedoch hier kurz aufgeführt werden, da insbesondere diese Ansätze der Führungsforschung derzeit als besonders vielversprechend behandelt werden.

Demnach korrelieren wesentliche Aspekte transformationaler Führung in den meisten Untersuchungen positiv mit der Mitarbeiterzufriedenheit und Arbeitsleistung, allerdings besteht Unklarheit über die zugrunde liegenden Prozesse, die bei Vorgesetztem und Mitarbeitern stattfinden. Außerdem bleiben die konkreten Verhaltensweisen ungeklärt, die situativen Variablen werden nicht spezifiziert und es wird auch nicht angegeben, welche nachteiligen Effekte transformationale Führung haben könnte, insbesondere, wenn etwa mehrere Führungskräfte mit unterschiedlichen Vorstellungen bei denselben Mitarbeitern solche Führungstechniken praktizieren.

Es beeinträchtigt den Geltungsbereich weiter, dass transformationale Führung lediglich dyadische Beziehungen zwischen Vorgesetztem und Mitarbeiter berücksichtigt, aber komplexere soziale Beziehungen und Vorgänge wie etwa die Beziehungen zwischen den Mitarbeitern ausschließt. Schließlich scheint es, als ob hier der Einfluss von Vorgesetzten auf ihre Mitarbeiter, vielleicht als Folge von Wunschdenken, erheblich überschätzt wird.

Mit ähnlichen Argumenten lassen sich auch die charismatischen Führungstheorien in Frage stellen. Hinzu kommt, dass Charisma phänomenologisch inkonsistent beschrieben wird, und unklar bleibt, unter welchen Bedingungen Charisma gewonnen oder verloren wird, wobei die Theorien annehmen, dass beides stattfinden kann. Aufgrund der vielen Unklarheiten plädiert zum Beispiel Wofford (1999) für kontrollierte Experimente zu diesem Zweck. Eine positivere Sichtweise vertritt Rosenstiel (2001), der betont, „dass derartige Ansätze keinesfalls einen Rückfall in die simplen monokausalen Erklärungsansätze der klassischen personalistischen Führungstheorie darstellen“ (Seite 330). Vielmehr seien sie als eine Erweiterung der interaktionistischen Kontingenztheorien anzusehen.

3.4.5 Subjektive Führungstheorien

Subjektive Führungstheorien (Biedermann, 1989) sind auch unter den Begriffen implizite Eignungstheorien, naive Führungstheorien oder unter anderen Bezeichnungen bekannt. Dabei handelt es sich um Ansätze, die sich damit beschäftigen, welche Vorstellungen Führungskräfte oder andere Personengruppen über typische oder ideale Eigenschaften von Führungskräften haben, die sich vor allem auf eigenschaftsbezogene Personenmerkmale konzentrieren. Um dies herauszufinden, werden üblicherweise mittels offener oder strukturierter Befragungstechniken die Vorstellungen der Befragten ermittelt. Die Ergebnisse werden mit quantitativen oder qualitativen Auswertungsverfahren mehr oder weniger differenziert und verdichtet. Neben den Eigenschaften von Führungskräften werden manchmal die Bedingungen und Konsequenzen dieser Eigenschaften erfasst und thematisiert (Schilling, 2001).

Nicht immer entspricht das Vorgehen den Standards wissenschaftlicher psychologischer Untersuchungen, was Erhebungs- und Auswertungsmethode sowie theoretische Fundierung betrifft. Häufig werden Diskussionsbeiträge in Workshops, deren Zustandekommen ebenfalls nicht immer geklärt ist, anhand nicht offen gelegter Methoden zusammengefasst. Außerdem gibt es viele Untersuchungen, die selbst gar nicht explizit subjektive Führungstheorien erfassen wollen, obwohl sie tatsächlich genau dies tun (Schieffer, 1998). Stattdessen beanspruchen sie, auf diese Weise objektive Eignungsanforderungen an Führungskräfte zu erhalten, wie sie dann beispielsweise die Grundlage von Verfahren zur Personalauswahl und -entwicklung bilden.

Häufig resultieren subjektive Führungstheorien daraus, wenn als Vorbereitung für managementdiagnostische Verfahren eine Anforderungsanalyse durchgeführt wird, indem so genannte Experten einschätzen, welche Persönlichkeits- und Fähigkeitseigenschaften führungsrelevant sind. Umso geringer die Befragung strukturiert ist, desto eher kann man davon ausgehen, dass damit eher ein subjektives Bild der Experten erfasst wird, statt realer Anforderungen (Kannheiser, 1995). Das ist vor allem dann problematisch, wenn diese Ergebnisse nicht explizit als subjektive Persönlichkeitstheorien verwendet werden, sondern einen allgemein gültigen Anspruch erheben, hiermit relevante Eignungsaspekte als Basis für die Eignungsdiagnostik zu liefern.

3.5 Fazit

Wie weit liefern nun diese Ansätze brauchbare Informationen über Anforderungen einer Führungsposition und deren Diagnostik? Funktionale Analysen stellen einen viel versprechenden Zugang dar, weil sie am klarsten formulie-

ren, worin die Anforderungen an eine Person als Führungskraft bestehen, auch wenn sie auf notwendige Prozesse in einer Organisation fokussieren, und damit die Bedeutung des Einzelnen scheinbar ausklammern. Aktivitätsstudien können ebenfalls nützliche Informationen zu diesem Zweck liefern, wobei ihre Ergebnisse gelegentlich auf der Ebene konkreter Aktivitäten wie Telefonieren verhaftet bleiben, ohne diese in einen psychologischen oder organisatorischen Kontext einzuordnen. Allerdings haben sie eine empirische Basis.

Auch Kategoriensysteme und Taxonomien erscheinen interessant, da sie als Instrumente zur Erfassung oder Beschreibung von Führungsinhalten verwendet werden und als solche inhaltlichen Kriterien psychologischer Theorien und methodischen Ansprüchen genügen müssen. Integrative Ansätze haben schließlich den Vorteil, dass sie durch ihren Rückgriff auf verschiedene Arbeiten führungsrelevante Aspekte breit abdecken.

Traittheorien können ihre Aussagen über die Relevanz bestimmter Eigenschaftsbündel für Führungserfolg nicht immer schlüssig begründen oder belegen. Ähnliches gilt für die teilweise präskriptiven Aussagen von Verhaltenstheorien, der Führungsstilforschung sowie der situativen und Kontingenztheorien. Von den transformationalen und den charismatischen Führungstheorien steht noch eine Konkretisierung der postulierten Prozesse aus. Mit den angesprochenen psychischen Prozessen werden kognitive Ansätze wiederum interessant. Indirekt weisen systemische Führungstheorien auf die Relevanz von Problemlösefähigkeit und ähnlichen Eigenschaften hin.

Kapitel 4

Führungsrelevante Eigenschaften?

4.1 Übersicht

Einen Schwerpunkt der Outcome- oder Ziel-Evaluation sens. Chen (1990) bilden die Spezifizierung der Ziele der Diagnostik. In der Managementdiagnostik sind diese meistens in Form von Anforderungs- oder Kompetenzmodellen definiert, welche das diagnostische Instrumentarium abdecken soll. Traittheorien der Führung wollen die besonderen Eigenschaften oder Eigenschaftsbündel identifizieren, mit denen sich Führungskräfte von anderen Personen abheben. Dieser Ansatz wurde im Abschnitt 3.3.1 bereits diskutiert. Im vorliegenden Kapitel wird anhand der Befundlage diskutiert, welche Fähigkeiten und Persönlichkeitseigenschaften im Einzelnen für den beruflichen Erfolg von Führungskräften bedeutsam sind.

Die Diskussion gliedert sich nach leistungsbezogenen Eigenschaften, die vor allem kognitive Fähigkeiten ansprechen, und nach nicht-leistungsbezogenen Eigenschaften, die eher auf individuelle Präferenzen zielen. Am Ende der Erörterungen steht als Vorschlag für eine Outcome-Evaluation ein Anforderungsmodell, das Problemlöseprozesse und interpersonelle Anforderungen in den Mittelpunkt stellt.

4.2 Leistungsbezogene Eigenschaften

Ob und wie weit leistungsbezogene Eigenschaften den beruflichen Erfolg von Führungskräften bestimmen, diskutieren Wissenschaftler ebenso kontrovers wie Praktiker. Das hängt zum einen vom jeweiligen Verständnis von Führung ab, das kognitive, soziale und emotionale Aspekte unterschiedlich stark gewichtet, zum anderen davon, welche Vorstellung über den Inhalt der Arbeitstätigkeit von Führungskräften besteht. Traditionellerweise wird Führung

als ein interpersonelles Phänomen angesehen, das sich in den Interaktionen zwischen Führungskräften und Mitarbeitern zeigt. Der Begriff Management hebt mehr auf Prozesse wie etwa Planung, Koordination oder Entscheiden ab, womit eher kognitive Anforderungen angesprochen sind.

Typischerweise wird von einer Führungskraft beides gefordert. Wie der Arbeitsdefinition von Yukl (2002) bereits zu entnehmen ist, beinhaltet eine Führungsposition nicht allein die Aufgabe, Mitarbeiter anzuleiten. Außerdem umfasst sie weitere Funktionen, beispielsweise Strategien auszuarbeiten oder umzusetzen, was konzeptuell eher den kognitiv leistungsbezogenen Bereich als den sozialen Bereich anspricht. Ob die Fähigkeit zur Bewältigung sozialer Situationen eine andere als intellektuelle Leistungsfähigkeit ist, wird gelegentlich in Frage gestellt. Vielmehr wird zunehmend die Bedeutung von kognitiven Prozessen als Querschnittsfunktion für die Bewältigung verschiedener, auch sozialer Situationen betont. Im Mittelpunkt der Diskussion kognitiver Prozesse steht in dieser Arbeit das komplexe Problemlösen, das im Vergleich mit anderen, möglicherweise relevanten leistungsbezogenen Eigenschaften und Prozessen erörtert wird (Kickul & Neumann, 2000).

4.2.1 Intelligenz

Es erscheint plausibel, dass aufgrund der großen Verantwortung und der Vielfalt der Informationen und Aufgaben, die von Führungskräften zu bewältigen sind, eine besonders hohe kognitive Leistungsfähigkeit erwartet wird. Verschiedene Intelligenzmodelle und deren psychometrische Erfassung sind in der einschlägigen Literatur bereits ausführlich dargestellt. Sternberg und Kaufmann (1998) charakterisieren in einem Übersichtsartikel dazu die Vielfalt der Zugänge zur Intelligenzforschung mit dem Kommentar: „The study of intelligence is like a real-world jeopardy game.“

Die Studien zum Zusammenhang von Intelligenz mit beruflichem Erfolg unterstützen mehrheitlich solche Plausibilitätsannahmen, wie zum Beispiel der Literaturüberblick über Verfahren der Personalauswahl von Borman, Hanson und Hedge (1997), die zusammenfassende Darstellung von Cook (1998) oder der Übersichtsartikel von Hough & Oswald (2000). Zumeist berichten diese Untersuchungen substantielle, wenn auch nicht immer befriedigende Korrelationen. Dabei ist hervorzuheben, dass Intelligenz von allen bisher untersuchten Prädiktoren für alle Berufsgruppen den höchsten Zusammenhang mit beruflichem Erfolg aufweist (Hunter & Hunter, 1984).

Dennoch wird die Intelligenzdiagnostik häufig kritisiert. Ein Kritikpunkt ist die sogenannte Schwellenhypothese, wonach nur ein minimales Fähigkeitsniveau für gute Arbeitsleistung gegeben sein muss, und ober- oder unterhalb dieses Niveaus keine weiteren Unterschiede bezüglich der Arbeitsleistung be-

stehen (Wallach, 1976). Diese Annahme spiegelt sich für die Managementdiagnostik beispielsweise in Matarazzos (1972) Behauptung wider, bei Führungskräften, Wissenschaftlern und generell Akademikern sei kaum ein Intelligenzquotient unter 115 zu finden. Allerdings war es schwierig, diesen oder einen anderen Cutoff zu belegen. Für die unterschiedlichsten Tätigkeiten, konnte Varianz in der beruflichen Leistung durch Intelligenzunterschiede aufgeklärt werden. Ein weiterer Einwand besagt, dass Intelligenz ein zu allgemeines Konstrukt sei, um speziellen Erfordernissen einer bestimmten Tätigkeit wie einer Führungsposition gerecht zu werden. Auch dies konnte empirisch widerlegt werden (Hunter & Hunter, 1984).

	Allgem. Intelligenz	Wahrn.- geschw.	Psycho- motorik	Aggreg. Maß
Manager	.53	.43	.26	.53
Schreibkraft	.54	.46	.29	.55
Verkauf	.61	.40	.29	.62
Sicherheit	.42	.37	.26	.43
Service	.48	.20	.27	.49
Handel	.46	.43	.34	.50
Fertigung	.37	.37	.40	.47
Autobau	.28	.31	.44	.46
Verkäufer	.27	.22	.17	.28

Tabelle 4.1: Korrelationen zwischen Leistungsmaßen und Berufserfolg bei verschiedenen Berufsgruppen (Metaanalyse von Hunter & Hunter, 1984).

Andere Kritikpunkte beziehen sich auf die zu geringe Korrelation der Intelligenzwerte mit beruflicher Bewährung, die insgesamt sehr unterschiedlich ausfällt, aber sich verschiedenen Metaanalysen zufolge mehrheitlich maximal um $r=0.30$ bewegt (Ghiselli, 1973). Diese Werte wurden allgemein als unbefriedigend empfunden, was innerhalb der allgemeinen Krise der Persönlichkeitspsychologie zu dieser Zeit den Erwartungen entsprach. Jedoch kamen Schmidt und Hunter (1977) zu dem Schluss, dass aus verschiedenen methodischen Gründen die Korrelationen unterschätzt wurden.

So sind die meist kleinen Stichproben besonders anfällig für Stichprobenfehler, was die große Varianz der Befunde erklärte. In vielen Studien wurde nur eine bereits vorselektierte Personenstichprobe untersucht, nämlich die erfolgreichen Bewerber. Darüber hinaus begrenzte die Reliabilität von Prädiktor und Kriterium ebenfalls die maximal mögliche Korrelation. Deshalb schätzten Schmidt und Hunter (1977) die so genannte wahre Validität, indem sie die Validitätskoeffizienten um die Effekte der eingeschränkten Reliabilität und der eingeschränkten Varianz korrigierten. Die so berechnete Validität ist nach einer Faustregel etwa doppelt so hoch wie der ursprüngliche Wert.

Intelligenztests liefern einerseits Information über die allgemeine Intelligenz als aggregiertem Wert, das so genannte g , das bereits einen Großteil der Varianz des Kriteriums Berufserfolg aufklärt. Zum anderen geben die Teilskalen je nach zugrunde liegendem Intelligenzmodell die Ausprägung spezifischer Fähigkeiten an. Letztere konnten meist keine inkrementelle Varianzaufklärung über den Beitrag der allgemeinen Intelligenz hinaus leisten (Larson & Wolfe, 1995). Ob dies tatsächlich eindeutig dafür spricht, dass die Vorhersage von beruflicher Bewährung „not much more than g “ sei (Ree, Earles & Teachout, 1994), ist fraglich.

Zunächst könnte man auf der Ebene eines allgemeinen Intelligenzbegriffs damit argumentieren, dass generell Intelligenz entscheidend ist für die erfolgreiche Bewältigung kritischer Situationen. Dies macht die Befundlage deutlich, ohne Intelligenzdefinitionen im Detail zu diskutieren (Robertson & Smith, 2001). Auch für Führungskräfte ist es entscheidend, kritische Situationen zu meistern. Weiterhin könnte man im Rahmen eines gewählten Intelligenzstrukturmodells genauer bestimmen, welche Komponenten davon wie etwa Gedächtnis, verbale oder numerische Fähigkeiten im Führungsalltag gefordert sind. Um dies zu untermauern, wäre eine Arbeitsanalyse erforderlich sowie eine genauere Modellierung der Arbeitsleistung auf Kriterienseite. Beides hat bisher kaum stattgefunden.

Vielmehr kann gemeinsam mit Borman, Hanson und Hedge (1997) vermutet werden, dass allgemeine Informationen auf Prädiktorseite ebenso allgemeinen Kriterien gegenübergestellt wurden, was aufgrund der gleichen Aggregationsebene angemessen ist (Brunswik, 1955). Jedoch sollten speziellen Prädiktoren wie spezifischen Intelligenzkomponenten auch spezielle Kriterien beruflicher Eignung gegenübergestellt werden. Unter diesen Umständen ist eine bessere Vorhersage durch die Teilskalen möglich, wenn auch nicht garantiert. Daneben ist zu überprüfen, ob sich durch die Hinzunahme weiterer berufsspezifischer Fähigkeiten inkrementelle Varianz aufklären lässt, wie zum Beispiel durch die Fähigkeit zum komplexen Problemlösen oder spezielle unternehmerische oder führungsbezogene Eigenschaften.

Ein weiteres Problem auf praktischer Seite ist die angeblich geringe Akzeptanz oder soziale Validität der gängigen Intelligenztests durch die Entscheidungsträger ebenso wie durch die Teilnehmer. Aus eignungsdiagnostischer Sicht sollte das nicht verhindern, dass Intelligenztests zur Personalauswahl eingesetzt werden, um so mehr, da die Testwerte am höchsten von allen einschlägigen Auswahlverfahren mit dem beruflichen Erfolg korrelieren (Cook, 1998). Außerdem ist die soziale Validität von Intelligenztest auf Bewerberseite nicht so gering, wie man vermutet hat, sondern das Verfahren wird vielmehr als fair empfunden (Schuler, 1990).

Dennoch werden Intelligenztests in der Personalauswahl nicht so häufig

eingesetzt, wie man vermuten könnte. Oft scheitert es an der ablehnenden Entscheidung des Betriebsrates, der bei internen Maßnahmen mit den verwendeten Instrumenten einverstanden sein muss. Festzuhalten bleibt, dass Intelligenzmaße offensichtlich berufliche Eignung in den meisten Bereichen mit substanziellen Beträgen prognostizieren können, also auch den Erfolg in Führungs- oder Managementaufgaben. Deshalb stellt Intelligenz einen wichtigen Aspekt führungsrelevanter Fähigkeiten dar.

4.2.2 Wissen

Zunächst trivial erscheint es, dass ein fundiertes Wissen über die eigene Arbeit notwendig ist, um diese erfolgreich ausführen zu können. Allerdings wird in der Praxis häufig wenig Wert darauf gelegt, dieses mit Hilfe diagnostischer Verfahren zu erfassen, da man davon ausgeht, dass der Bewerber seine fachliche Kompetenz durch Zeugnisse, formale Qualifikationen und seine Karrierelaufbahn bereits hinreichend bewiesen habe. Allgemein formuliert, stellt das Wissen eines Menschen seine interne Repräsentation der Realität dar. Er erwirbt und speichert es in der Auseinandersetzung mit seiner Umwelt (Kluwe, 1995a).

Dabei gibt es häufig zwei Formen von Wissen zu unterscheiden, die von Anderson (1982) begrifflich gefasst wurden. Demnach besteht Wissen einerseits aus einer deklarativen Komponente, also verbal beschreibbaren Gedächtnisinhalten über Sachverhalte, Vorgänge, Personen und Objekte. Daneben existiert eine prozedurale Wissenskomponente. Das sind Fertigkeiten, die eine Person verbal nicht beschreiben kann, obwohl sie diese beherrscht. Sie umfassen nicht nur motorische Abläufe, sondern auch bestimmte Strategien und Operationen beim Umgang mit Informationen oder Problemen. Damit rückt prozedurales Wissen in die Nähe des komplexen Problemlösens.

Alternativ zu den grundlagenwissenschaftlich orientierten Wissenstheorien entwickelten Wagner und Sternberg (1985) das Konzept des *Tacit knowledge*, das auf berufsspezifische Fähigkeiten ausgerichtet ist. Im Mittelpunkt des Tacit knowledge-Ansatzes steht das allgemeine praktische Handlungswissen. Es enthält die Fähigkeit, Alltagsprobleme zu lösen. Wagner und Sternberg setzen diesen Ansatz für verschiedene Berufsgruppen um. Bei Führungsaufgaben unterscheiden sie bei ihrem *Tacit knowledge for manager* drei Ebenen der praktischen Intelligenz, die bei Führungsaufgaben zusammenwirken.

Managing self bezieht sich auf die Maximierung der eigenen Produktivität und umfasst, verschiedene Aufgaben nach ihrer Wichtigkeit zu ordnen, effektive Mechanismen zur Aufgabenbewältigung zu kennen und Strategien zur Selbstmotivation zu beherrschen. Managing others bezieht sich in erster Linie auf Verhalten gegenüber Mitarbeitern. Dazu zählen Aufgabenverteilung,

Umgang mit Unzulänglichkeiten und Herstellung eines produktiven Betriebsklimas. Managing career beziehungsweise Managing tasks schließt Bereiche wie das Überzeugen von Kunden und das effektive Verteilen von Ressourcen ein. Mit dem TKIM (Tacit Knowledge Inventory for Managers), der die praktische Intelligenz in kaufmännischen Bereichen erfassen soll, steht ein Verfahren bereit, das die Messung dieser drei Formen praktischen Wissens anzielt. Es erfasst die Führungsexpertise auf drei Skalen, welche die Aspekte Managing self, Managing others und Managing self abdecken.

Ähnlich wie die Arbeiten von Wagner und Sternberg berichten Studien über die kognitive Leistung von Experten (Chi, Glaser & Farr, 1988), dass diese sowohl ein umfangreiches Faktenwissen als auch damit verknüpfte Lösestrategien nutzen, um fachspezifische Problemsituationen zu bewältigen. Auf der Grundlage dieser Überlegungen kann man zunächst davon ausgehen, dass Wissen ein führungsrelevantes Konstrukt ist. Dabei ist gleichzeitig unklar, in welcher Form es für berufseignungsdiagnostische Zwecke am besten erfasst wird, und welche Relevanz die einzelnen Wissensaspekte für beruflichen Erfolg haben.

Neuere Ansätze, die Führungsleistung modellieren, betrachten Wissen als eine erfolgsrelevante Komponente, wie etwa Connelly, Gilbert, Zaccaro, Threlfall, Marks und Mumford (2000). Im Unterschied zu anderen Auffassungen, die Wissen als redundanten Bestandteil von Problemlösefähigkeit betrachten, wird hier angenommen, dass Berufserfolg durch Wissen, durch Problemlösefähigkeit und durch soziale Urteilsfähigkeit bestimmt wird. Sofern sich das Wissen nicht auf allgemeine Strategien oder Inhalte bei der Arbeit bezieht, muss es für jede Situation spezifisch formuliert und durch Expertenbefragungen, Critical incident technique oder ähnliche Verfahren erfasst werden, woraus ein Instrument zur Überprüfung des Wissens bei den betreffenden Führungskräften erzeugt wird. Insofern sind Verfahren zur Erfassung des Wissens nicht verallgemeinerbar.

4.2.3 Komplexes Problemlösen

In der Literatur wird häufig behauptet, doch selten begründet, dass Anforderungen einer Managementaufgabe mit den Anforderungen korrespondieren, die aus der Bewältigung komplexer Probleme erwachsen, wie Kersting (1999) kritisiert. So kommt beispielsweise Putz-Osterloh (1995) zu dem Schluss, dass die Eigenschaften komplexer Probleme theoretisch mit den Situationen übereinstimmen, mit denen Führungskräfte typischerweise konfrontiert werden. Allerdings sieht sie dabei keine logische Äquivalenzbeziehung gegeben, vielmehr formuliert sie eine Implikation, wonach ein erfolgreicher Manager zwar kompetent sei, Probleme zu lösen, aber umgekehrt ein guter Problem-

löser nicht notwendigerweise eine gute Führungskraft sei. Es muss demzufolge managementspezifische Problemsituationen als Teilmenge aller denkbaren komplexen Probleme geben, die sie aber in ihrem Beitrag nicht spezifiziert, auch wenn sie auf verschiedene Eigenschaften von Problemen hinweist.

Diese Behauptung findet sich in ähnlicher Form in anderen Publikationen, etwa in etlichen Beiträgen der Übersichtswerke von Geilhardt und Mühlbradt (1995) oder Strauß und Kleinmann (1995). Kersting (1999) vermisst an einer solchen Position sowohl die theoretische Fundierung als auch den Rekurs auf entsprechende Anforderungsanalysen von Managementaufgaben und von computergestützten Szenarios, welche die Erfassung von Problemlösefähigkeit beanspruchen. Die hartnäckige Existenz und kontinuierliche Wiederholung dieser Annahme führt Kersting auf die Unbestimmtheit des Wortes *Problemlösen* zurück, das durch seinen großen Interpretationsspielraum auf beliebige Situationen angewandt werden könne.

Für das Konstrukt des komplexen Problemlösen gibt es in der Umgangssprache, in der betrieblichen Praxis ebenso wie in der Theorie viele Definitionen, wie man es bereits vom Intelligenzbegriff her kennt. Generell liegt für eine Person dann ein Problem vor, wenn eine aktuelle Situation in eine andere überführt werden soll, und die Person nicht weiß, wie sie es bewerkstelligen kann, wenn sie also eine Barriere überwinden muss. Damit werden Probleme von Aufgaben abgegrenzt, bei denen die Regeln zur Bewältigung bekannt sind. Weiterhin lassen sich Probleme danach einteilen, ob die möglichen Maßnahmen, der Ausgangszustand oder der Zielzustand bekannt sind.

Legt man den von Dörner (1976) eingeführten Problemlösebegriff zugrunde, ist die Situation bei einem komplexen Problem durch Komplexität, Intransparenz, Dynamik und Vernetztheit gekennzeichnet. Komplexität bedeutet, dass sehr viele Aspekte bei der Problemlösung zu berücksichtigen sind. Intransparenz ist ein Problem, wenn der Problemlöser es anhand der vorliegenden Informationen nicht vollständig durchschauen kann. Dynamik führt dazu, dass sich Zustände der Situation ohne Eingreifen verändern. Bei Vernetztheit bestehen zwischen Variablen gleichzeitig mehrere Beziehungen.

Problematisch an diesen Definitionen ist, dass sie nicht eindeutig ermöglichen, verschieden komplexe Probleme zu unterscheiden. So ist ungeklärt, ob alle vier Eigenschaften notwendige Bedingungen dafür darstellen, dass ein komplexes Problem gegeben ist, oder ob einzelne davon als hinreichende Bedingung genügen. Alleine der Begriff der Komplexität ist selbst aus systemtheoretischer Perspektive nicht befriedigend geklärt (Casti, 1979), dazu kommen in den psychologischen Auffassungen noch weitere Aspekte, die sich auf die Wahrnehmung der Situation durch den Problemlöser beziehen.

Grenzt man jedoch diese definitorischen Schwierigkeiten aus, kann auf der Basis der Problemlösesituation eine Fähigkeit zum komplexen Problemlösen

konzeptualisiert werden. Diese wird in Anlehnung an ein Anforderungsmodell von Dörner (1986) begriffen, der die Problemlöseleistung in Form des Gesamtergebnisses auf vier Anforderungskategorien zurückführt, die sich aus den charakteristischen Eigenschaften der Problemsituation ergeben. Dörners Ziel ist, mit der Beschreibung dieser Anforderungsdimensionen eine Basis zur Erfassung der Problemlösefähigkeit als Eigenschaftskonstrukt zu liefern. Daraus könnte man umgekehrt auf Eigenschaften einer Problemsituation rückschließen, die so beschaffen sein muss, dass sie diese Anforderungen an den Problemlöser stellt. Angesichts der Ambiguität, der Führungskräfte nach Neuberger (2002) häufig ausgesetzt sind, könnte dies eine relevante Anforderung darstellen.

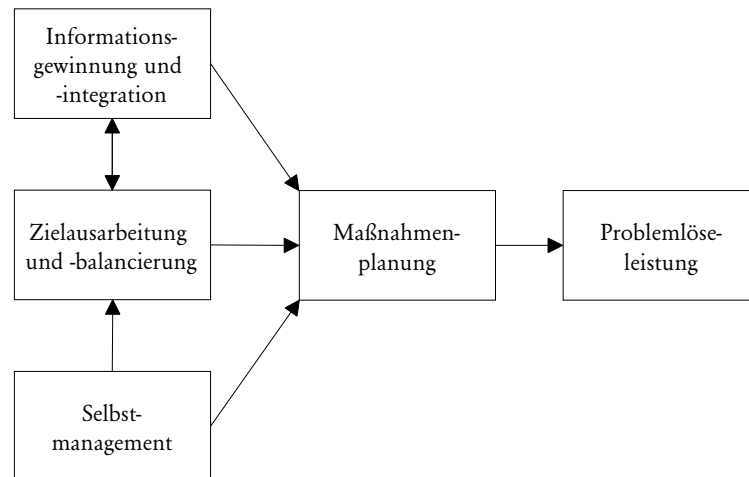


Abbildung 4.1: Modifiziertes Anforderungsmodell an Problemlöser, angelehnt an Dörner (1986).

Damit lässt sich die Fähigkeit zum komplexen Problemlösen als das Ausmaß, in dem die beschriebenen Anforderungen erfolgreich bewältigt werden, erfassen und diagnostizieren. Das Anforderungsmodell von Dörner (1986) wurde modifiziert, indem die vier Anforderungskategorien in eine hierarchische Struktur gebracht wurden. Demzufolge wirken die Informationsgewinnung, die Zielausarbeitung und das Selbstmanagement auf die Maßnahmenplanung, von der direkt die Problemlöseleistung abhängt. Dieses Modell wurde im Rahmen des computergestützten Szenarios FSYS operationalisiert (Wagener, 1994) und anhand regressionsanalytischer Modelle erfolgreich empirisch überprüft (Wagener, 2001).

In den wenigen Untersuchungen zur Kriteriumsvalidität der Problemlösefähigkeit wurde der Prädiktor meist über allgemeine Maße der Steuerleistung

operationalisiert. Dabei ergaben sich unterschiedlich hohe Korrelationen der Steuerleistung oder anderen Problemlösegrößen mit Berufserfolgskriterien, wie zum Beispiel der Übersicht bei Hasselmann (1993) zu entnehmen ist oder wie Kersting (1999) berichtet. Diese Unterschiede können verschiedene Ursachen haben. Zum einen waren die Stichproben meist klein, zum anderen ist die Operationalisierung von Berufserfolgskriterien mit den gleichen Problemen behaftet wie die der Prädiktoren und damit eine potenzielle Ursache für geringere Korrelationen mit den Prädiktoren als maximal möglich sind (De Corte, 1999). Neuere Untersuchungen mit umfangreichen Stichproben von Führungskräften finden sich bei Zaccaro, Mumford, Connelly, Marks und Gilbert (2000), einige Befunde daraus zeigt Tabelle 4.2.

	Problemlöse- fähigkeit	Lösungs- güte	Soz. Urteils- vermögen
Schreibfertigkeit	.31	.31	.32
Verbale Intelligenz	.14	.24	.20
Leistungsmotivation	.07	.15	.08
Direktivität	.07	.11	.07
Gewissenhaftigkeit	.24	.26	.31
Führungsexpertise	.33	.21	.27
Vorgesetztenurteil	.33	.21	.27

Tabelle 4.2: Korrelationen von Problemlösekompetenz und sozialer Kompetenz mit anderen Maßen bei Führungskräften (Zaccaro et al. 2000).

Neben den begrifflichen Schwierigkeiten ist es nicht einfach, Problemlösefähigkeit diagnostisch zu erfassen, meist mit computergestützten Szenarios (vergleiche Abschnitt 2.4.6). Es ist ein Schwachpunkt der betreffenden Forschungsarbeiten, spricht aber nicht notwendigerweise gegen den Einsatz des Konstrukts, dass bisher die Verwendung des Problemlösekonstrukts für die Berufseignungs- und Managementdiagnostik nicht überzeugend begründet wurde (Kersting, 1999).

Ein weiterer Kritikpunkt am komplexen Problemlösen ist die Frage, wie stark es sich mit Intelligenz und Wissen überschneidet. Süß (1996) geht davon aus, dass Problemlösefähigkeit ein redundantes Konstrukt sei, da er zeigen konnte, dass sich die Leistung im Szenario Hemdenfabrik vollständig durch Wissens- und Intelligenzmaße erklären lässt. Jedoch sind diese Befunde an ein einziges Szenario gebunden, und die Generalisierung ist fraglich. Zudem ist nicht geklärt, worin ein komplexes Problem besteht und wie es zu operationalisieren ist. Offensichtlich gibt es Überschneidungsbereiche zur Intelligenz, die überwiegend positiv mit Berufserfolg, auch bei Führungskräften korrelierte, so dass möglicherweise Problemlösen relevant ist.

Angesichts der Unklarheit des Problemlösekonstrukts und seiner Operationalisierungen greift diese Arbeit auf das Modell von Dörner (1986) zurück, das die Fähigkeit zum Lösen komplexer Probleme in Form von operationalisierbaren Anforderungen an den Problemlöser beschreibt. Ob dies der einzig sinnvolle Zugang zu der Problematik ist, kann zur Diskussion gestellt werden. Zu diesem Zweck wurde das gewählte Verständnis von Problemlösen expliziert.

4.2.4 Weitere Konstrukte

Die im Folgenden aufgeführten psychologischen Konstrukte überschneiden sich begrifflich mit dem komplexen Problemlösen. Inwieweit sie tatsächlich dazu redundant sind, wurde empirisch noch kaum untersucht. Außerdem sind sie nicht per se eignungsdiagnostische Konstrukte mit einem differenzialpsychologischen Hintergrund, sondern wurden implizit oder explizit aus einer allgemeinspsychologischen Perspektive heraus entwickelt. Die meisten Ansätze dieser Art wurden bisher nicht für diagnostische Zwecke umgesetzt.

Entscheiden

Sowohl aus betriebswirtschaftlicher wie auch aus psychologischer Perspektive kann die Entscheidungsforschung in drei Bereiche untergliedert werden. Eine Entscheidung ist die Wahl zwischen verschiedenen Alternativen in Bezug auf ein oder mehrere Ziele. Die normative Entscheidungsforschung will komplexe Entscheidungsprobleme strukturieren, in einem formalen Modell abbilden und nach logischen Kriterien Handlungsempfehlungen ableiten. Dagegen befasst sich die deskriptive Entscheidungsforschung mit der Modellierung des tatsächlichen Entscheidungsverhaltens von Individuen, Gruppen und Organisationen. In der präskriptiven Entscheidungsforschung geht es darum, Empfehlungen für optimale Entscheidungsstrategien unter bestimmten Umständen zu geben.

Für die vorliegende Fragestellung, nämlich die Identifikation relevanter Konstrukte für die Managementdiagnostik, ist die deskriptive Entscheidungsforschung relevant. Generell entwickelt und prüft sie Annahmen zum Entscheidungsprozess, der nach Jungermann (1995) in fünf Schritte aufgeteilt werden kann. Der erste Schritt besteht in der Problemrepräsentation, das heißt in der Aufgliederung in Ziele, Handlungsoptionen und Konsequenzen. Im zweiten Schritt wird der Wert der Handlungsmöglichkeiten beurteilt, der sich aus subjektiven oder objektiven Nutzen und Kosten zusammensetzt. Die Abschätzung, wie wahrscheinlich denkbare Handlungskonsequenzen sind, erfolgt im dritten Schritt, was die Grundlage für den vierten Schritt, der Wahl

einer Handlungsalternative bildet. Welche Wahl dann realisiert wird, ist Gegenstand des fünften Schrittes.

Deskriptive Entscheidungstheorien unterscheiden sich darin, ob und wie sie die einzelnen Schritte modellieren (Bröder, 2000). Allerdings liefern sie wenig Grundlagen für eine diagnostische Beurteilung des Entscheidungsprozesses, der sich auf die Art der Entscheidungsfindung und das Entscheidungsergebnis beziehen kann. Außerdem fehlen für die Managementdiagnostik Untersuchungen charakteristischer Entscheidungssituationen für Führungskräfte sowie Kriterien für die Qualität solcher Entscheidungen.

Planung, Organisation und Kontrolle

Weitere Konzepte, die als äußerliche Merkmale die intellektuellen Anforderungen im Alltag einer Führungskraft widerspiegeln, sind Planung, Organisation und Kontrolle. Dies entspricht der verkürzten Version der klassischen fünf Managementfunktionen nach Fayol (1929), die Management als institutionalisiertes Handeln auffassen. Planen wird als vorweggenommenes Handeln oder mentales Probehandeln verstanden. Der Strategiebegriff verweist auf längerfristige, globale Planung, mit Taktik ist dagegen kurzfristige lokale Planung und Reaktion gemeint. Planung, Organisation und Kontrolle vermitteln zwischen externalem Managementhandeln und internalem Probehandeln.

Nach Staehle (1999) setzt sich Management aus verschiedenen, aufeinander bezogenen Funktionsbereichen zusammen, die auch als Managementzyklus oder Managementprozess bezeichnet werden. Die Zielsetzung beinhaltet, Unternehmens- und Abteilungsziele zu entwickeln und vorzugeben. Bei der Alternativensuche geht es darum, verschiedene mögliche Aktivitäten zur Zielerreichung zu identifizieren. Aufgaben und Tätigkeiten werden bei der Entscheidung zielbezogen ausgewählt und festgelegt.

Mit der Planung ist gemeint, aufgabenbezogene Aktivitäten im Hinblick auf die Entscheidung zu koordinieren und zu delegieren. Um Tätigkeiten planungsbezogen umzusetzen, ist Organisation erforderlich, deren Resultate im Rahmen der Kontrolle ziel- und planungsorientiert beurteilt werden. Diesen Funktionen sind in der Praxis entwickelte Arbeitstechniken zugeordnet, die durch die verschiedenen Management-by-Konzepte beschrieben werden.

In diesem Zusammenhang spricht für die Relevanz dieser Konzepte, dass Planung, Organisation und Kontrolle nicht nur kognitive Operationen darstellen, sondern auch Führungs- oder Managementfunktionen. Allerdings lassen sich die damit verbundenen Prozesse in das gewählte Problemlösemodell einordnen, weshalb sie hier nicht als einzelne und eigenständige Konstrukte einbezogen werden.

Lernfähigkeit

Lernen wird gemeinhin als Verhaltensänderung aufgrund von Erfahrungen definiert. Sofern man davon ausgehen kann, dass Management ein Prozess ist, der eine Umgebung gestaltet und aufrechterhält, in der Ziele effizient erreicht werden können, muss eine Führungskraft in der Lage sein, auf Veränderungen zu reagieren. Das bildet sich in den Anforderungskriterien vieler Assessment-Center ab, die Lernfähigkeit als wesentliches Kriterium enthalten. Im Unterschied zu anderen Persönlichkeitskonstrukten zielt das Konzept der Lernfähigkeit auf intraindividuelle Leistungsunterschiede ab, die sich jedoch interindividuell unterscheiden (Guthke, 1982).

Demzufolge ist Lernfähigkeit eine Leistungsdisposition, die es dem Individuum ermöglicht, sich kognitiv mit der Umwelt auseinanderzusetzen. Guthke operationalisiert dies in seinen Lerntests, wobei Probanden einen intellektuellen Leistungstest bekommen, der nach einer Trainingsphase in ähnlicher Form erneut angewandt wird. Indikatoren für die Lernfähigkeit sind der Differenzwert zwischen den beiden Tests sowie das Abschneiden in der zweiten Durchführung. Vermutlich aufgrund des großen Aufwandes wird dieses Verfahren noch kaum in der beruflichen Eignungsdiagnostik praktiziert.

Im Unterschied dazu bleibt bei den Assessment-Centern, die dies explizit erfassen wollen, unklar, wie die Lernfähigkeit überhaupt operationalisiert ist. Im Lernpotenzial-AC nach Sarges (1995) werden die Kandidaten aufgefordert, sich vor dem eigentlichen Assessment einschlägig zu informieren. Aus den interindividuellen Unterschieden der späteren Performanz wird auf den Lernerfolg und damit auf die Lernfähigkeit geschlossen, die hierbei jedoch nur durch die interindividuellen, nicht die intraindividuellen Unterschiede repräsentiert wird.

Auf diese Weise wird also nicht, wie im ursprünglichen Lerntestkonzept vorgesehen, die Differenz zwischen der Leistung vor und nach einer Lerneinheit erfasst, worin die eigentliche Lernfähigkeit gesehen wird. Aus den interindividuellen Differenzen hier kann nur auf einen Wissensstand geschlossen werden, aber nicht darauf, wie groß der Zuwachs durch die Vorbereitung ist. Ob man unter diesen Umständen von einem Lernpotenzial-Assessment sprechen kann, ist fraglich. Darüber hinaus ist in dem wissenschaftlich fragwürdigen Begriff des Potenzials der Anspruch enthalten, eine Aussage über zukünftige Entwicklungen zu treffen.

Außerdem ist Lernen als Anpassung an veränderte Umstände durch Informationsverarbeitung oder durch Verwertung von Feedback nach Handlungen im gewählten Problemlösemodell enthalten. Es ist sicherlich für Führungskräfte und in vielen anderen Bereichen relevant, hier soll es nicht als eigenständiges Konzept verwendet werden.

4.3 Nicht leistungsbezogene Eigenschaften

Inwieweit nicht leistungsbezogene Eigenschaften generell Relevanz für beruflichen Erfolg haben, und insbesondere für Führungskräfte, bleibt strittig. Gleich, ob diese Eigenschaften über Fragebogen, im Einzelgespräch oder im Assessment-Center eingeschätzt werden, liegt ihre prognostische Validität hinter den leistungsbezogenen Eigenschaften zurück. Die Gründe dafür können in den verwendeten Verfahren liegen, die Reaktionen im Sinne sozialer Erwünschtheit erleichtern oder generell methodische Mängel aufweisen.

Möglicherweise sind diese Eigenschaften für den beruflichen Erfolg nicht so entscheidend, wie in der Praxis häufig angenommen. Betrachtet man einige der nicht leistungsbezogenen Eigenschaften wie interpersonelle Fertigkeiten oder Kommunikationsfähigkeit genauer, so decken sie auch Aspekte kognitiver Leistungsfähigkeit ab, wie zum Beispiel verbale Intelligenz. Warum Verfahren zur Erfassung dieser Persönlichkeitseigenschaften dennoch in der betrieblichen Eignungsdiagnostik eingesetzt werden, ist schwer nachzuvollziehen, da sogar die soziale Validität oder Akzeptanz von Leistungs- und Fähigkeitstests höher ist als die von Persönlichkeitstests (Schuler, 1990).

Ein Vorteil der tatsächlich oder vordergründig nicht leistungsrelevanten Konstrukte liegt vielleicht in der Außenkommunikation, da mit einem Wert auf einer Persönlichkeitsskala zunächst keine Wertung verbunden ist, wie es bei einem niedrigen Intelligenztestwert der Fall wäre. Daneben berichten Schuler, Frier und Kaufmann (1993), dass generell die Einsatzhäufigkeit eines Verfahrens höher mit der Akzeptanz bei Anwendern und Betroffenen korreliert als mit der Praktikabilität und Validität.

4.3.1 Persönlichkeit

Neben den in traittheoretischen Führungsansätzen (Abschnitt 3.3.1) postulierten Persönlichkeitsbündeln können auch einzelne Persönlichkeitseigenschaften und persönlichkeitspsychologische Modelle diskutiert werden. Die Traitansätze zur Führung zielen dabei insbesondere auf Leistungsmotivation, Machtmotivation und soziale Fertigkeiten. Welche Persönlichkeitseigenschaften besonders für Management-Aufgaben relevant sind, und welche Ausprägung sie haben sollen, wird meist anhand empirischer Validitätsuntersuchungen geprüft. Viele neuere Untersuchungen gehen von dem Big-Five-Modell der Persönlichkeit aus (McCrae & Costa, 1987), oder es werden bei Metaanalysen die Befunde in diesem Sinne zusammengefasst.

Dieses fünffaktorielle Modell unterscheidet Extraversion, Emotionalität, soziale Verträglichkeit, Gewissenhaftigkeit und Aufgeschlossenheit. Barick und Mount (1991) zeigten in ihrer Metaanalyse, dass Extraversion für Mana-

ger und Verkäufer den Berufserfolg vorhersagen kann und Gewissenhaftigkeit in allen Berufsgruppen wichtig ist. Allerdings beträgt der Validitätskoeffizient selbst nach Korrektur bei Extraversion $r=0.18$ und bei Gewissenhaftigkeit liegt er zwischen $r=0.20$ und $r=0.23$, wobei diese beiden Persönlichkeitseigenschaften einen signifikanten inkrementellen Beitrag zu Leistungstest lieferten. Alle weiteren Persönlichkeitsdimensionen leisteten nach dieser Studie keine bedeutsame Vorhersage beruflichen Erfolgs. Neuere Befunde einer großen internationalen Studie an Führungskräften (Boudreau, Boswell & Judge, 2001) zeigt Tabelle 4.3 Hier korrelieren Neurotizismus und Verträglichkeit am höchsten (negativ) mit verschiedenen Berufserfolgskriterien.

	Gehalt	Aufstieg	Verantwortung	Vielseitigkeit
Neurotizismus	-.30	-.21	-.10	-.03
Extraversion	.04	.08	.10	.14
Aufgeschlossenheit	-.02	-.03	.01	.03
Verträglichkeit	-.24	-.14	-.14	-.15
Gewissenhaftigkeit	-.07	-.05	-.01	-.04

Tabelle 4.3: Korrelationen von NEO-FFI-Skalen mit Berufserfolg bei Führungskräften (Boudreau, Boswell & Judge, 2001).

Insgesamt erscheint die Befundlage zum Zusammenhang von Persönlichkeit und Berufserfolg uneinheitlicher als bei den Leistungsmaßen. Bis Mitte der neunziger Jahre wurde der Wert von Persönlichkeitstests oder -skalen für die Berufseignungsdiagnostik, insbesondere für die Personalauswahl, eher skeptisch bewertet (Hough, 1992; Tett, Jackson, Rothstein & Reddon, 1994) und kontrovers diskutiert, sowohl in Primäruntersuchungen wie in Metaanalysen (Ones, Mount, Barick & Hunter, 1994).

Ähnlich wie bei der Diskussion von Intelligenzmaßen geht es hier um methodische Aspekte der Metaanalyse, also um die Frage, welche Studien einzubeziehen sind, ob und wie eine Varianzkorrektur der Validitätskoeffizienten stattfindet oder wie das Berufserfolgskriterium beschaffen ist. Darüber hinaus stellt sich bei der Metaanalyse der Kriteriumsvalidität von Persönlichkeitsmaßen die Frage, ob das Vorzeichen der Korrelation berücksichtigt oder ignoriert werden soll, da sowohl negative wie positive Korrelationen für die Brauchbarkeit sprechen. Außerdem hängt die Validität stärker von dem jeweiligen Instrument ab, als dies bei Leistungstests der Fall ist.

Weitere Punkte betreffen die Breite oder Enge von Persönlichkeitsmaßen in Bezug auf das Kriterium, die möglichst übereinstimmen sollten (Ones & Viswesvaran, 1996; Paunonen, Rothstein & Jackson, 1999). Bei spezifischen Kriterien sollten nach Paunonen et al. auch die Prädiktoren so beschaffen sein, was am ehesten vor dem Hintergrund einer Tätigkeitsanalyse erreicht

werden kann. Eine sehr optimistische Einschätzung vertreten neuere Arbeiten, wie etwa Goodstein und Lanyon (1999), welche die Brauchbarkeit von Persönlichkeitsmaßen für die Personalauswahl ebenso wie die Personalentwicklung betonen. Demzufolge können sie insbesondere im Führungsbereich wertvolle Beiträge leisten.

Diese und andere Beiträge führen selten eine vergleichende Diskussion mit Leistungsmaßen. Sofern dies doch geschieht, wird den Leistungsmaßen zur Last gelegt, dass sie im Unterschied zu Persönlichkeitsmaßen oder alternativen Prädiktoren Minderheiten benachteiligen. Dies geht allerdings auf Kosten der Kriteriumsvalidität (Hoffmann & Thornton, 1997). Zudem lässt sich ein Interaktionseffekt zwischen Persönlichkeit und intellektueller Leistungsfähigkeit beziehungsweise eine inkrementelle Varianzaufklärung durch Persönlichkeitsmaße nicht zuverlässig zeigen, obwohl der Zusammenhang plausibel erscheint (Sackett, Gruys & Ellingson, 1998).

Es erschwert eine vergleichende und integrative Diskussion weiter, dass Studien und Metaanalysen die Interkorrelationen zwischen Persönlichkeit und Leistungsfähigkeit sowie weiteren Prädiktoren von Berufserfolg nicht genügend berücksichtigen. Doch legen einige Arbeiten die Vermutung nahe, dass die berichteten Korrelationen zwischen Persönlichkeit und beruflicher Leistung durch die Interkorrelation zwischen Persönlichkeits- und Leistungsmaßen erklärbar sind (Bobko, Roth & Potosky, 1999). Daneben besteht nach wie vor das Problem, dass die Resultate von Persönlichkeitstests von den Probanden manipuliert werden können (Borman, Hanson & Hedge, 1997). In dieser Arbeit wird eine eigene Metaanalyse zu Persönlichkeitseigenschaften und Führungserfolg durchgeführt.

4.3.2 Motivation

„Eine von der Psychologie bisher nicht hinreichend beantwortete Frage ist diejenige, was ein Motiv (beziehungsweise was Motivation) sei“ (Herrmann, 1991). Die Vielfalt der Definitionsversuche lässt sich darauf reduzieren, dass Motivation zu den intervenierenden Variablen gehört, also nicht beobachtet werden kann, aber aus dem Verhalten erschließbar ist. Motive sind damit Teildeterminanten des Verhaltens, die außerdem durch das Ziel charakterisiert sind, auf welches das Verhalten ausgerichtet wird. Allgemein wird unter Motivation eine generelle Anstrengungsbereitschaft oder Leistungsmotivation verstanden. Diese spielt in der Berufseignungsdiagnostik eine wichtige Rolle.

Motivationstheorien können danach eingeteilt werden, worin sie Motive verursacht sehen. McClelland (1955) unterscheidet vier Motivationsmodelle. Überlebensmodelle sehen als Quelle der Motive biologisch definierte Bedürf-

nisse, die durch ihre Befriedigung reduziert werden. In Reizstärkemodellen wird die Neigung zum Reagieren oder Verhalten als abhängig vom Ausmaß des auslösenden Reizes angesehen. Ähnliche Argumente enthält das Reizmustermodell, das Motive als zentralnervöse Zustände des Individuums sieht, die durch die Beziehungen äußerer Reize bedingt sind. Nach den Annahmen des Modells der Gefühlsauslösung sind Motive erlernt, die in Diskrepanzen zwischen den Erwartungen und Wahrnehmungen des Individuums bestehen.

Eine andere Möglichkeit, Motivationstheorien zu gruppieren, ist die Einteilung nach den Motivationsinhalten. Für managementdiagnostische Zwecke erscheint zunächst die Leistungsmotivation relevant. Sie wird definiert als eine Persönlichkeitsdisposition, die in Situationen angeregt wird, in denen es zu einer Auseinandersetzung mit Gütemaßstäben kommt. Dabei ist das Erleben und Verhalten auf die Zielereignisse Erfolg beziehungsweise Misserfolg ausgerichtet, die angestrebt oder vermieden werden sollen.

Demnach werden Personen von dem Motiv Hoffnung auf Erfolg oder Vermeidung von Misserfolg angetrieben. Erfolgsmotivierte Personen neigen dazu, in Leistungssituationen relevante Informationen effizienter auszuwerten, sich realistischere Ziele zu setzen sowie ausdauernder und erfolgreicher zu arbeiten (Heckhausen, Schmalt & Schneider, 1985). Bisher gibt es wenige Untersuchungen zur prädiktiven Validität von Leistungsmotivation für den Managementbereich, die außerdem zu unterschiedlichen Ergebnissen über die Bedeutung von Leistungsmotivation im Zusammenhang mit Machtmotivation kommen (Wainer & Rubin, 1971).

Schuler & Prochaska (2000) beschreiben die Entwicklung und Konstruktvalidierung des Leistungsmotivationsinventars (LMI). Dem Verfahren liegt ein weites Verständnis der Leistungsmotivation zugrunde, die auf 17 Skalen erfasst wird und eine differenzierte Beschreibung ermöglichen soll. Im Rahmen der Konstruktvalidierung wurden weitere Charakteristika und Außenkriterien von Personen erhoben. Die engsten Beziehungen bestanden zu den Big-Five-Persönlichkeitsskalen Neurotizismus und Gewissenhaftigkeit, aber auch zur Übernahme von Führungsaufgaben.

Aus dem Bereich der traittheoretischen Führungstheorien kommt das Konzept der Führungsmotivation, dessen Definition vom Verständnis des Führungsbegriffs abhängt. Dementsprechend existieren dazu sehr heterogene Ansätze. Zunächst kann Führungsmotivation einfach als Motiv zur sozialen Einflussnahme aufgefasst werden, oder als Machtmotiv. Weitere Ansätze konzipieren Führungsmotivation als zeit- und situationsstabile Motivkonstellation, wie etwa McClelland und Boyatzis (1982), die eine Kombination aus hoher Machtmotivation, geringer Anschlussmotivation und starker Selbstdisziplin als ideale Voraussetzung für Führungserfolg ansehen, was empirisch nicht belegt werden konnte.

Miner (1990) verknüpft die Motivationsstruktur mit organisationalen Gegebenheiten, wobei nur ideale Kombinationen in Führungserfolg resultieren. Auch hier sind die empirischen Belege ebenso wenig überzeugend wie die Argumentation (Nachreiner & Müller, 1995). Ebenso wie bei der Leistungsmotivation ist ein wesentliches Problem dieser Konzepte die Messung über projektive Verfahren, die von vornherein mit eingeschränkter Objektivität und Reliabilität belastet sind.

Im Zusammenhang mit Motivation können auch Wertorientierung sowie Loyalität diskutiert werden, womit die Bereitschaft gemeint ist, sich mit dem Unternehmen zu identifizieren und sich die Unternehmensziele zu eigen zu machen. Allerdings wird dies bei der Personaleignungsdiagnostik in der Regel nicht explizit diagnostiziert, sondern implizit aus sekundären Informationen erschlossen, aber dennoch bewertet (Rosenstiel, 1995).

Im Zusammenhang mit Motivation ist außerdem der Aspekt der Interessen zu diskutieren. Interesse ist ein persönlicher, positiv gefärbter Bezug zu bestimmten Zielbereichen, mit denen sich eine Person besonders häufig und intensiv auseinander setzt, und der ein bewusster Bestandteil des Selbstkonzeptes ist. Dabei kann zwischen Interesse als Eigenschaft einer Person und einer beruflichen Aufgabe unterschieden werden, die einander für optimale Aufgabenerfüllung entsprechen müssen. Dies umzusetzen, wurde für die Managementdiagnostik wiederholt gefordert, stagnierte aber angesichts der widersprüchlichen Ergebnisse aus der Berufsinteressenforschung sowie der daraus erwachsenden Zweifel an der Prognosetauglichkeit des Konzeptes (Haeberlin, 1995).

Ähnlich konnten Sackett, Gruys und Ellington (1998) durch Leistungsmotivation keine signifikante inkrementelle Varianzaufklärung der Vorhersage von beruflicher Leistung nachweisen, wenn daneben Maße intellektueller Leistungsfähigkeit verwendet wurden. Als Fazit zur Motivation als führungsrelevante Eigenschaft bleibt, dass einige Ansätze mit gewisser Plausibilität bestehen, aber die Operationalisierung und die Befundlage problematisch erscheint, ebenso wie die Übertragung von der allgemeinpsychologischen Grundlagenforschung auf Fragestellungen der Managementdiagnostik erst in Ansätzen erfolgt ist.

4.3.3 Interpersonelle Fertigkeiten

In der Praxis wird immer wieder betont, dass vor allem Führungskräfte soziale oder auch kommunikative Kompetenz benötigen, die als eine so genannte Schlüsselqualifikation angesehen wird. Allgemein versteht man darunter die außerfachliche und nur mittelbar leistungsbezogene Fähigkeit, andere Menschen, zu verstehen und angemessen mit ihnen umzugehen. So detailliert

wie Intelligenz ist es als eignungsdiagnostisches Konstrukt noch nicht erschlossen, jedoch sollen im Folgenden einige Ansätze berichtet werden. Wissenschaftssprachlich wird soziale Intelligenz von der akademischen Intelligenz unterschieden.

Thorndike (1920) führte den Begriff der sozialen Intelligenz in die Intelligenzforschung ein und definierte ihn als Fähigkeit, andere zu verstehen, zu leiten, und in persönlichen Beziehungen weise zu agieren. Mit ihrer Verwendung in den verschiedenen Gebieten wie betrieblicher Eignungsdiagnostik, klinischer Psychologie und Intelligenzforschung hat die soziale Kompetenz einen großen Bedeutungsspielraum mit unterschiedlichen Schwerpunkten. So kann sie als Kognition aufgefasst werden, die zum Verständnis der Hintergründe des Verhaltens anderer führt, oder als soziale Empathiefähigkeit auf emotionaler Seite. Als weiterer Indikator wird gelegentlich die Teilnahme an Gemeinschaftsprozessen betont oder in den Vordergrund gestellt. Daneben soll die soziale Intelligenz im Sinne von Empathie die Rollenübernahme erleichtern, welche handlungswirksame Kommunikationsstrategien zur Durchsetzung und günstigen Selbstdarstellung ermöglicht.

Die Erfassung sozialer Intelligenz ist ein Schwerpunkt von Assessment-Centern, wobei die Operationalisierung hierbei ebenso uneinheitlich erscheint wie die Verwendung und Definition des Begriffs. Daneben existieren Tests zur Erfassung sozialer Intelligenz, die Bilder und schriftliche Beschreibung interpersoneller Probleme als Reizmaterial einsetzen. Die Vineland Social Maturity Scale (VSMS) von Doll (1953) sieht in der Sozialkompetenz die Befähigung des Individuums, sich selbst helfen zu können und mit anderen Kontakt aufzunehmen. Damit ist die Verantwortung für sich und andere Menschen verbunden. Doll schlägt die Bildung eines Sozialquotienten vor, wobei analog zur frühen Intelligenzdiagnostik das Sozialalter durch das Lebensalter dividiert wird. Dieses Instrument scheint jedoch keine Fortentwicklung erfahren zu haben. Neuere diagnostische Instrumente zur Erfassung sozialer Kompetenz findet man in der wirtschaftspsychologischen Forschung selten, natürlich gibt es Ausnahmen, wie etwa der Beitrag von Funke und Schuler (1998).

Weiterhin lässt sich die soziale Kompetenz, wenn sie in sozialer Umgänglichkeit besteht, mit Persönlichkeitsskalen wie Extraversion, Offenheit oder Verträglichkeit erfassen. Ein Instrument, das diese Aspekte sozialer Kompetenz beinhaltet, wäre der Fragebogen zur Erfassung des unternehmerischen Potentials von Müller (1999) mit den Skalen Leistungsmotivation, Kontrollbedürfnis, Risikobereitschaft, Problemorientierung und Präferenz sozialer Einflussnahme. Generell gestaltet sich die reliable und valide Erfassung der sozialen Intelligenz als Konstrukt aber schwierig, was eine kontinuierliche Forschung behinderte (Walker & Foley, 1973).

Am ehesten besteht soziale Kompetenz, auch als interpersonelle Kompetenz bezeichnet, wohl in einer Kommunikations- oder Artikulationsfähigkeit, die sich treffender über verbale Intelligenz charakterisieren lässt. Nach Weinstein (1969) umfasst sie die Fähigkeit, Aufgaben im Umgang mit anderen Personen zu bewältigen und beinhaltet außerdem eine wertfrei gemeinte Manipulation des Verhaltens anderer zugunsten eigener, psychotherapeutischer oder sozialisierender Zwecke.

Ob diese Fähigkeit über eng umgrenzte konkrete Situationen hinaus generalisierbar ist, blieb bis heute ungeklärt. Die Selbstbehauptung oder Selbstsicherheit spielt vor allem in der Verhaltenstherapie eine wichtige Rolle und wird als Ziel bei der Behandlung sozialer Phobien angestrebt. Wer über diese Eigenschaft verfügt, kann in sozialen Konflikten adäquat reagieren sowie seine eigenen Bedürfnisse artikulieren und durchsetzen, ohne dabei andere zu beeinträchtigen. Nach Zimmer (1978) ist sie als situationsübergreifendes Persönlichkeitsmerkmal empirisch nicht bestätigt.

Soziale Intelligenz im Sinne einer Fähigkeit zur Einschätzung sozialer Situationen konnte durch allgemeine Intelligenz weitgehend aufgeklärt werden. So korrelieren in einer Untersuchung von Stevens und Campion (1994) die Testergebnisse zur Teamfähigkeit zu $r=0.81$ mit allgemeiner Intelligenz. Damit ist die konzeptuell geforderte Abgrenzung zur akademischen Intelligenz empirisch nicht nachweisbar, ebenso wie die gängige Behauptung, dass eine hohe Ausprägung akademischer Intelligenz mit einer geringeren Ausprägung sozialer Fertigkeiten einhergehe oder umgekehrt.

Nur wenige Arbeiten bemühen sich um eine wissenschaftliche Definition des Begriffs sowie eine Operationalisierung. McFall (1982) argumentiert ausgehend von der psychometrischen Erfassung von sozialer Kompetenz. Diese manifestiert sich in der Beobachtung des Verhaltens in kritischen Situationen, ist aber durch nicht beobachtbare *social skills* determiniert. Diese bestehen aus Fähigkeiten zur Dekodierung von Informationen in einer Situation, Entscheidungsfähigkeit, und Kommunikationsfähigkeit. Karkoschka (1998) entwickelt eine Arbeitsdefinition sozialer Kompetenz als Basis einer eignungsdiagnostischen Erfassung, allerdings nicht speziell im Führungskräftebereich. Demzufolge orientieren sich sozial kompetente Menschen in sozialen Interaktionen an ihren Zielen und setzen zur Zielerreichung ihre verfügbaren Mittel der Situation angemessen ein, womit sich soziale Kompetenz nur situationsspezifisch äußern kann.

Schneider, Ackerman und Kanfer (1997) fassen soziale Kompetenz ähnlich als zusammengesetzte Variable. Diese besteht aus den Komponenten Einsicht in soziale Vorgänge, soziale Anpassung, angemessenes Verhalten in sozialen Situationen, Aufgeschlossenheit gegenüber anderen, Einflussnahme, Wärme und Extraversion. Dabei gehen sie von einer Überschneidung mit Persön-

lichkeitsvariablen in Anlehnung an ein Big-Five-Konzept aus sowie von einer Überschneidung mit kognitiven Fähigkeiten vor allem im verbalen Bereich. Sie beschreiben die Eigenschaften von Situationen, in denen diese kognitive Komponente besonders zum Tragen kommt. Das ist vor allem dann der Fall, wenn diese Situationen neuartig sind, wenn sie soziale Informationen enthalten, die nicht mit eigenen Erwartungen übereinstimmen und wenn sie so strukturiert sind, dass das eigene Verhalten impliziten sozialen Normen entsprechen soll, die in der Situation entdeckt werden müssen.

Die Vielzahl der Konzepte zu sozialer Intelligenz oder sozialer Kompetenz macht eine psychometrische Umsetzung für die Managementdiagnostik ebenso problematisch wie der mangelnde Beleg des Zusammenhangs mit beruflichem Erfolg. Im Widerspruch zu den Grundannahmen des Konzepts erscheint es sogar vielversprechender, die soziale Intelligenz über Maße der allgemeinen Intelligenz oder über andere kognitive Prozesse zu erfassen, denen man eine Querschnittsfunktion für soziale und nicht-soziale Probleme unterstellen kann (Fleishman et al., 1991). Trotzdem sollten noch weitere Bemühungen investiert werden, um die Diskrepanz zwischen allgemeiner Auffassung und Befundlage aufzuklären oder zu beheben.

Viele Beiträge über Aufgaben und Anforderungen an Führungskräfte zielen auf interpersonelle Aspekte von Führung und Management. Möglicherweise sind zwar die Anforderungen, die sich auf soziale Situationen beziehen, ähnlich strukturiert wie die in nicht-sozialen Situationen, man könnte wie Fleishman et al. (1991) beides in eine Querschnittsfunktion integrieren. Solange nicht geklärt ist, ob und wie interpersonelle und intellektuelle Fähigkeiten zusammenhängen, sollten die interpersonellen Aspekte von Führung als ein eigenständiges Konstrukt in die weiteren Überlegungen mit einbezogen und auf der Grundlage der Führungsliteratur erweitert werden, die auf soziale Prozesse immer wieder hinweist.

4.4 Anforderungsmodell

Diese erste Betrachtung demonstrierte, dass es viele ähnliche theoretische Ansätze gibt, die sich vielleicht integrieren lassen. Ein Problem der verschiedenen Konstrukte besteht darin, dass sie begrifflich und empirisch nicht überschneidungsfrei sind. Die Abgrenzung zwischen Intelligenz und Problemlösefähigkeit, zwischen Intelligenz und Lernfähigkeit, zwischen Problemlösefähigkeit und Entscheiden beziehungsweise Planung, Organisation und Kontrolle ist deshalb schwierig vorzunehmen. Am umfassendsten erscheint zu diesem Zweck das beschriebene Anforderungsmodell komplexer Probleme, das um den interpersonellen Bereich erweitert wird (Tabelle 4.4).

Kategorie	Beschreibung
Informationsmanagement Informationsgewinnung Informationsintegration Planungsinformation Situationsinformation Kontrollinformation Systeminformation	Informationsgewinnung, -Integration Strategien zur Informationsbeschaffung Verarbeitung von Informationen Zur Vorbereitung von Maßnahmen Situationszusammenhänge Resultate eigener Maßnahmen Systemzustände
Zielbildung Zielausarbeitung Zielableitung Zielgewichtung	Zielausarbeitung und -balancierung Eigene Festlegung von Zielen Ableitung von Teilzielen Gewichtung von Teilzielen
Maßnahmenplanung Bündelung Prioritätensetzung Effektivität Entscheidungsverhalten Anordnung Steuerung	Beschluss von Maßnahmen Bündelung von Maßnahmen Festlegung der zeitlichen Reihenfolge Intensität von Maßnahmen Risikokalkulation Verabschiedung der Maßnahme Korrektur oder Vermeiden von Fehlern
Selbstmanagement Frustrationstoleranz Belastbarkeit Wertkonflikte	Emotionale Stabilität Umgang mit Misserfolgen Umgang mit Stressoren Konflikte mit eigenem Wertesystem
Interpersonelle Anforderungen Kooperation Einflussnahme Konfliktmanagement Kommunikation Durchsetzung Rücksicht	Interaktion mit anderen Zusammenarbeit mit anderen Beeinflussung anderer Adäquates Verhalten in Konflikten Artikulation eigener Position Eigene Zieldurchsetzung Berücksichtigung der Ziele anderer

Tabelle 4.4: Vorschlag für ein problemlösezentriertes Anforderungsmodell bei Führungskräften.

Die Aspekte Denken und Planen, Entscheiden und strategisch Handeln fielen heraus, da sie konzeptuell im Rahmen der Problemlösefähigkeit gefasst werden können und außerdem wenig Anhaltspunkte zur Diagnostik geben können. Auf weitere Konstrukte wie Extraversion und Gewissenhaftigkeit wird ebenfalls verzichtet, weil diese ebenso wie andere Eigenschaften nicht direkt in Bezug zu den verhaltens- und anforderungsorientierten Arbeiten zu bringen sind. Motivation wäre sicher interessant, und besticht durch hohe Plausibilität, aber die Operationalisierung und Befundlage dazu erschien im Verhältnis dazu eher enttäuschend.

4.4.1 Hauptkategorien Problemlösen

Erwartet wird hier, dass eine Führungstätigkeit vor allem verschiedene Aspekte der Fähigkeit zum komplexen Problemlösen enthält, was sich im Kategoriensystem niederschlagen soll. Nach Dörner (1986) stellt ein komplexes Problem vier grundlegende Anforderungen an den Problemlöser, die hier als Hauptkategorien gefasst werden. Dieses Problemlösemodell wurde bereits operationalisiert, in seiner internen Struktur geprüft, weiterentwickelt und validiert (Wagener, 2001). Für die Konkretisierung dieser von Dörner eher allgemein gehaltenen Dimensionen in Teilkategorien werden diese zur Operationalisierung entwickelten Ansätze einbezogen.

Informationsgewinnung und -integration als erste Hauptkategorie steht im Zusammenhang mit der Intransparenz komplexer Probleme. Sie liegt dann vor, wenn der Problemlöser die Situation nicht vollständig erfasst hat, was typischerweise den Ausgangszustand darstellt. Dies erfordert Aktivitäten zur Beschaffung und sinnvollen Verwertung dieser notwendigen Informationen über das Problem, über Zusammenhänge der Situation, über Maßnahmen und deren Wirkung sowie über den Zustand kritischer Variablen. Darin enthalten sind also Aspekte der Orientierung und Kontrolle sowie Strategien der Informationsbeschaffung als Teilkategorien.

1. Beschaffung von Information
2. Verarbeitung von Information
3. Information zur Planung von Maßnahmen
4. Orientierung über Situationszusammenhänge
5. Information über die Resultate eigener Maßnahmen (Kontrolle)
6. Information über den Zustand kritischer Variablen

Eine weitere Anforderung ist die Hauptkategorie Zielausarbeitung und -balancierung, wenn Ziele vage oder global vorgegeben sind. Als Teilkategorien und Einzelanforderungen müssen Teilziele abgeleitet und konkretisiert werden, die unter Umständen im Widerspruch zueinander stehen, weshalb des Weiteren eine entsprechende Gewichtung der Ziele, eine Zielbalancierung erforderlich wird, die bis zum Verzicht auf Teilziele führen kann.

1. Eigene Zielausarbeitung
2. Ableitung von Teilzielen aus einem globalen Ziel
3. Gewichtung von Teilzielen

Unmittelbar relevant für den Problemlöseerfolg ist die Maßnahmenplanung und -anordnung als dritte Hauptkategorie mit sechs Teilkategorien. Sobald die Situation und die Ziele bekannt sind, müssen Maßnahmen geplant und angeordnet werden, die in der Regel ein Bündel von Einzelmaßnahmen beinhalten, die sich auf verschiedene Eingriffsvariablen bezüglich des Zieles beziehen. Hierbei geht es meist auch nicht allein um die Anordnung einer punktuellen Maßnahme oder eines solchen Bündels, sondern die Maßnahmen müssen in eine Folge gebracht werden, wobei spätere Maßnahmen auf die Resultate vorhergegangener Maßnahmen aufbauen.

1. Koordination der Maßnahmen
2. Priorisierung der Maßnahmen bezüglich zeitlicher Abfolge
3. Festlegung angemessener Intensität der Maßnahme
4. Risikokalkulation der alternativen Maßnahmen
5. Veranlassung der Maßnahme
6. Vermeidung von Fehlsteuerung

Das Selbstmanagement als letzte Hauptkategorie des Problemlösens bezeichnet die Fähigkeit, trotz Stressoren die emotionale Verfassung soweit stabil zu halten, dass die eigenen Fähigkeiten und Gewohnheiten der Situation angepasst werden können. Die Eigenschaften, die eine Situation als komplexes Problem charakterisieren, führen zu Zeitdruck, Unsicherheit über die Wirkung der angeordneten Maßnahmen und damit zu Stress. Eventuelle Misserfolge bei der Steuerung erzeugen Frustration. Möglicherweise muss man Maßnahmen anordnen, die dem eigenen Wertesystem widersprechen, wie Lohnkürzungen oder Entlassungen in einem Betrieb. Deshalb ist emotionale Belastbarkeit im Umgang mit diesen Faktoren erforderlich.

1. Umgang mit Frustration
2. Umgang mit Stressoren
3. Umgang mit eigenen Wertkonflikten

4.4.2 Hauptkategorie interpersoneller Bereich

Nicht so eindeutig lässt sich entscheiden, auf welches Modell man im interpersonellen Bereich zurückgreifen soll. Eine denkbare Basis wäre die Definition interpersonaler Kompetenz von Weinstein (1969), die hier in der Fähigkeit besteht, Aufgaben im Umgang mit anderen Personen zu bewältigen und das Verhalten anderer beeinflussen zu können. Ergänzend können die Ansätze

zur Selbstbehauptung in der klinischen Psychologie genutzt werden. Nimmt man die Sicht der Führungsliteratur dazu, so kann man interpersonelle Anforderungen, die sich also im Zusammenhang mit anderen Personen stellen, auf fünf Bereiche verdichten. So ist die Zusammenarbeit mit anderen Personen ein wichtiger Punkt. Insbesondere als Führungskraft geht es dabei häufig um deren Beeinflussung. Treten Konflikte auf, ist ein adäquates Verhalten nötig, um diese möglichst konstruktiv zu behandeln. Außerdem muss eine Führungskraft in der Lage sein, ihre eigenen Anliegen klar zu kommunizieren. Die Durchsetzung eigener Ziele muss häufig gelingen. Ebenso wichtig ist, dass die Ziele und Bedürfnisse anderer berücksichtigt werden.

Vor diesem Hintergrund werden die verschiedenen Ansätze zum interpersonellen Bereich speziell im Bezug auf Führung verdichtet. Damit enthält die Hauptkategorie der interpersonellen Anforderungen als Teilkategorien:

1. Zusammenarbeit mit anderen Personen
2. Beeinflussung anderer Personen
3. Adäquates Verhalten in Konflikten
4. Artikulation eigener Position
5. Durchsetzung eigener Ziele
6. Berücksichtigung der Ziele anderer

4.5 Fazit

Es wurde diskutiert, wie brauchbar verschiedene gängige Eigenschaftskonstrukte für die Managementdiagnostik sein könnten. Davon ausgehend wurde ein Anforderungsmodell vorgeschlagen, das sich auf ein diagnostisches Problemlösemodell stützt sowie auf interpersonelle Anforderungen. Es soll im Folgenden die Grundlage für zwei Studien zur Outcome-Evaluation sowie eine Impact-Evaluation bilden. Eine Studie beinhaltet die Entwicklung eines Kompetenzmodells in der Praxis, die zweite eine Impact-Evaluation entlang diesen Ansatzes. Dort ermöglicht es eine theoriegeleitete Bewertung der empirischen Befunde. Schließlich wird mit einer literaturgestützten Anforderungsanalyse ein Forschungsmodell zur Outcome-Evaluation vorgestellt.

Kapitel 5

Outcome-Evaluation: Praxisdaten

5.1 Übersicht

Die bisher vorgestellte Literaturlage weist darauf hin, dass sich Anforderungen in Führungsaufgaben unter anderem als kognitive Eignungskonstrukte modellieren ließen, mit Schwerpunkt auf Problemlösefähigkeit. Auch wenn es andere Möglichkeiten gibt, liegt ein Fokus dieser Arbeit auf diesem Problemlösemodell als Eignungskonstrukt, ergänzt um interpersonelle Anforderungen. Bevor dies in den folgenden Kapiteln an zwei empirischen Evaluationsstudien und einer systematischen Literaturanalyse umgesetzt wird, sollen hier konkretere empirische Informationen dargestellt und ausgewertet werden, welche die Perspektive der personaldiagnostischen Praxis und Anwendungsforschung widerspiegeln.

Immer wenn Anwendungsforschung betrieben wird, ist es im Sinne einer Outcome-Evaluation (Chen, 1990) wichtig, bei einer Maßnahme zunächst festzustellen, welche Ziele die *stakeholder* oder Interessenträger verfolgen. Dabei sind mehrere Fragen zu beantworten: Sind die Ziele konsistent? Lassen sie sich als messbare Kriterien operationalisieren? Sind sie aus einer fachlich-wissenschaftlichen Perspektive her zu begründen? Chen weist darauf hin, dass einerseits die Ziele der Interessenträger aus der Praxis einbezogen werden müssen, um die Akzeptanz und Relevanz der Maßnahme und ihrer Ergebnisse zu optimieren. Andererseits sollte der Wissenschaftler, der die Maßnahme begleitet und die Evaluation durchführt, seine Fachkompetenz in die Zieldefinition und Gestaltung der Maßnahme einbringen, weil er in der Regel über mehr Informationen verfügt.

Deshalb behandelt dieses Kapitel zunächst die Frage der Anwenderperspektive aus drei empirischen Studien. Ziel der Maßnahmen ist ja, diese Anforderungen möglichst gut abzubilden. Die systematische qualitative Aus-

wertung von Stellenanzeigen zeigt, welche allgemeinen fachlichen und nicht-fachlichen Anforderungen an Nachwuchsführungskräfte dort formuliert werden, und welche Rolle kognitive und persönlichkeitsbezogene Eigenschaften spielen. Wie in einem eigenen Beratungsprojekt die Anwenderbedürfnisse stufenweise mit wissenschaftlich begründeten Vorschlägen in einem Kompetenzmodell für obere Führungskräfte integriert wurde, das einer Potenzial-einschätzung zugrundegelegt wurde, zeigt die zweite Studie. In der dritten Studie wurden schließlich neuere Validierungsstudien zu Persönlichkeitskonstrukten im Rahmen des Big-Five-Ansatzes bei Führungskräften metaanalytisch zusammengefasst. Dies greift zwar auf wissenschaftliche Literatur zurück, reflektiert aber dennoch die Anwenderperspektive, da die Untersuchungen im Praxisfeld und an der Zielgruppe Führungskräfte stattfanden.

5.2 Auswertung von Stellenanzeigen

Im äußeren Kontext dieser Arbeit wurde immer wieder die Frage aufgeworfen, welches implizite Eignungsmodell (Biedermann, 1989) die Praktiker aus ihrer Personalarbeit als bewährt ansehen. Es ist plausibel, dass neben theoretischen Analysen und wissenschaftlichen Befunden auch dieses einbezogen werden muss, und es sollte bei der Entwicklung eignungsdiagnostischer Konzepte und Maßnahmen berücksichtigt werden. So wird verhindert, dass wissenschaftliche Konzepte und praktische Anwendungen zusammenhangslos entwickelt beziehungsweise umgesetzt werden. Wo finden sich also praktische Konzepte, und wie kann man diese zusammenfassen? Aktuellere Daten über die Situation in Deutschland könnten ermittelt werden, indem man beispielsweise die Selbstdarstellungen der Unternehmen gegenüber Bewerbern auswertet.

5.2.1 Methode

Als Quelle für eine Sammlung dieser Stellenanzeigen wurde der *Stellenreport* für Hochschulabsolventen herangezogen. Er wendet sich an Hochschulen und Fachhochschulen und enthält Firmeninformationen zusammen mit Stellenanzeigen. Dabei handelt es sich um eine Kurzübersicht über Unternehmen, die an Führungsnachwuchs interessiert sind. Neben den angebotenen Tätigkeitsbereichen und gewünschten Anforderungen geben sie, anders als die Stellenanzeigen in Zeitungen, an, mit welchen Auswahlverfahren die Bewerber zu rechnen haben. Die Unternehmen präsentieren sich mit Kontaktadresse und weiteren Informationen, wobei es sich um größere und renommierte Firmen handelt.

Kategorie	Beschreibung
Fachliche Qualifikationen	
Praxiserfahrung	Berufliche Erfahrung
Abschlussnote	Examensnote mindestens gut oder sehr gut
Sprachkenntnisse	Besondere Fremdsprachenkenntnisse
Studienrichtung	Benennung gewünschter Fachrichtungen
Auslandserfahrung	Auslandsstudium, -praktika oder -arbeit
Studiendauer	Zügiges Studium innerhalb Regelstudienzeit
Außerfachliche Kompetenzen	
Flexibilität	Mobilität, Flexibilität
Engagement	Initiative, Verantwortungsbereitschaft
Teamfähigkeit	Zusammenarbeit mit anderen
Sozialkompetenz	Soziale Kompetenz, Kommunikation
Persönlichkeit	Persönlichkeit
Analytische Fähigkeiten	Analytisches, konzeptionelles Denken
Eingesetzte Verfahren	
Unterlagen	Auswertung der Bewerbungsunterlagen
Interview	Bewerbungsgespräch, Einstellungsinterview
Assessment Center	Assessment Center
Sonstiges	Planspiel, Fallstudie, Persönlichkeitstest

Tabelle 5.1: Kategorien bei der Auswertung von Stellenanzeigen.

Von insgesamt 83 deutschen oder internationalen Unternehmen mit deutscher Niederlassung wurden für die vorliegende Arbeit fachliche sowie nichtfachliche Einstellungskriterien und die zu erwartenden Verfahren erfasst und ausgezählt. Zu diesem Zweck wurde eine qualitative Inhaltsanalyse durchgeführt, die zunächst eine zusammenfassende Inhaltsanalyse vornahm, was eine inhaltliche Strukturierung des Materials ermöglichte, die schließlich ausgezählt und deskriptiv ausgewertet wurde. Damit wurde hier ein induktives Vorgehen gewählt, um möglichst gut das Material abzubilden, und dieses mit dem theoriegeleiteten Anforderungsmodell zu kontrastieren.

So entstand ein Kategoriensystem mit drei Hauptkategorien (Tabelle 5.1). Die ersten beiden enthalten die fachlichen und nichtfachlichen Anforderungen mit jeweils sechs Teilkategorien. Die dritte Hauptkategorie umfasst die Art der verwendeten Diagnoseverfahren mit vier Teilkategorien. Die Angaben zu den Verfahren waren so allgemein, so dass sich diese Hauptkategorie nicht weiter untergliedern ließ. Diese bestanden in Auswertung der Bewerbungsunterlagen, Gespräche beziehungsweise Interviews, Assessment-Center sowie weiteren Verfahren wie etwa Fallstudien, Problemlöseaufgaben oder ähnliche, die sich nicht in die anderen drei Kategorien einordnen ließen, aber nur wenige Male genannt waren.

Bei den Eignungskriterien enthielten die fachlichen Anforderungen kon-

krete, belegbare Qualifikationen. Darunter fallen die Teilkategorien, Studienrichtung, kurze Studiendauer, eine gute Abschlussnote, praktische Erfahrungen, bestimmte Sprachkenntnisse sowie Auslandsaufenthalte. Außerfachliche Kompetenzen sind dadurch gekennzeichnet, dass sie Persönlichkeitsbereich anzusiedeln sind und sich nicht durch Zertifikate, Zeugnisse oder ähnliches nachweisen lassen. Dies sind für die Unternehmen Teamfähigkeit, soziale oder kommunikative Kompetenz, Flexibilität und Mobilität, Engagement beziehungsweise Initiative oder Verantwortung, analytische oder konzeptionelle Fähigkeiten sowie einfach Persönlichkeit.

Die quantitative Auswertung wurde bewusst auf eine deskriptive Häufigkeitsanalyse beschränkt, weil das Ziel hier in einer Status-quo-Aufnahme kommunizierter Anforderungen besteht. Aufgrund der Stichprobengröße wären weitere inferenzstatistische Häufigkeitsvergleiche sowie Korrelationsauswertungen möglich. Darauf wurde verzichtet, weil dazu zum einen spezifische Hypothesen bestehen müssten und zum anderen durch die unterschiedlichen Branchen, Unternehmensgrößen und weitere Eigenschaften die Stichprobe so heterogen ist, dass diese Ergebnisse mehrfach konfundiert wären und sich kaum sinnvoll interpretieren ließen.

5.2.2 Ergebnisse

So wurde bei der Auswertung des *Stellenreports* (siehe Abschnitt 5.2.1) neben den Verfahren auch erfasst, welche Anforderungen die 83 Unternehmen an die potentiellen Bewerber formulieren. Die Anzeigen im *Stellenreport* waren als standardisierte, tabellarische Übersicht aufgebaut, die in jeweils gleich benannte Rubriken unterteilt waren. Welchen Inhalt die Rubrik Auswahlverfahren hatte, wurde oben bereits dargestellt. Die Inhalte der Rubrik Einstellungskriterien bestanden ebenfalls in stichwortartigen Aufzählungen, die sich auf wiederkehrende, sehr ähnliche oder gleiche Begriffe beschränkten, weshalb sie ebenfalls leicht auszuzählen waren.

Insgesamt zeigt die Auswertung, dass den Unternehmen fachliche Qualifikationen kaum wichtiger als außerfachliche Kompetenzen sind, da beide fast genauso häufig genannt wurden. Jedoch scheinen die angebotenen Positionen nicht durchgängig bestimmte fachliche oder außerfachliche Eignungskriterien zu erfordern, denn keines der Einstellungskriterien wurde in allen Positionen verlangt.

An fachlichen Qualifikationen waren einschlägige Praxiserfahrung am wichtigsten, außerdem die Abschlussnote sowie bestimmte Sprachkenntnisse. Da sich der *Stellenreport* explizit an Hochschulabsolventen wendet, ist bemerkenswert, dass Studierende aus der Sicht der Unternehmen bereits während des Studiums Praxiserfahrung erwerben sollen. Trotz der kontrovers geführ-

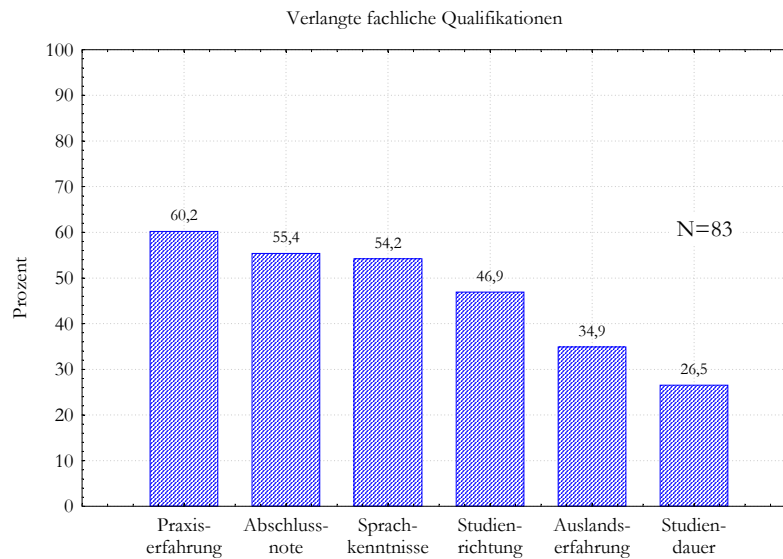


Abbildung 5.1: Anteil der Unternehmen, welche die jeweilige Qualifikation als Einstellungskriterium angaben (aus *Stellenreport 2000/2001*)

ten Diskussion um die Bedeutung von Abschlussnoten und Studiendauer stellen diese in der Praxis offensichtlich ein wichtiges Selektionskriterium dar.

Von den außerfachlichen Kompetenzen standen Flexibilität, Engagement und Teamfähigkeit im Vordergrund. Flexibilität trat in den Stellenanzeigen meist bezogen auf Mobilität auf, so dass damit wohl eher die Bereitschaft zum Ortswechsel als eine kognitive Eigenschaft gemeint ist. Engagement wurde in der Regel spezifiziert durch das Adjektiv außerfachlich oder außercurricular, weshalb dies wahrscheinlich keine grundsätzliche Arbeitshaltung im Sinne von Motivation, sondern die nachweisliche Teilnahme an entsprechenden Aktivitäten beinhaltet. Weniger wichtig waren soziale beziehungsweise kommunikative Kompetenzen, Persönlichkeit, und am seltensten erwähnten die Unternehmen analytische oder konzeptionelle Fähigkeiten. Unter Teamfähigkeit und sozialer Kompetenz scheinen die Unternehmen verschiedene Konzepte zu verstehen, da sie weder systematisch ausschließlich noch gemeinsam auftraten.

Auffällig ist bei den außerfachlichen Kompetenzen, dass ihre Bezeichnungen sehr allgemein gehalten sind. Bekanntlich sollte eine Stellenanzeige ihre Anforderungen möglichst präzise formulieren, damit sich potenzielle Bewerber im Vorfeld selbst selektieren können, was für beide Seiten wirtschaftlicher

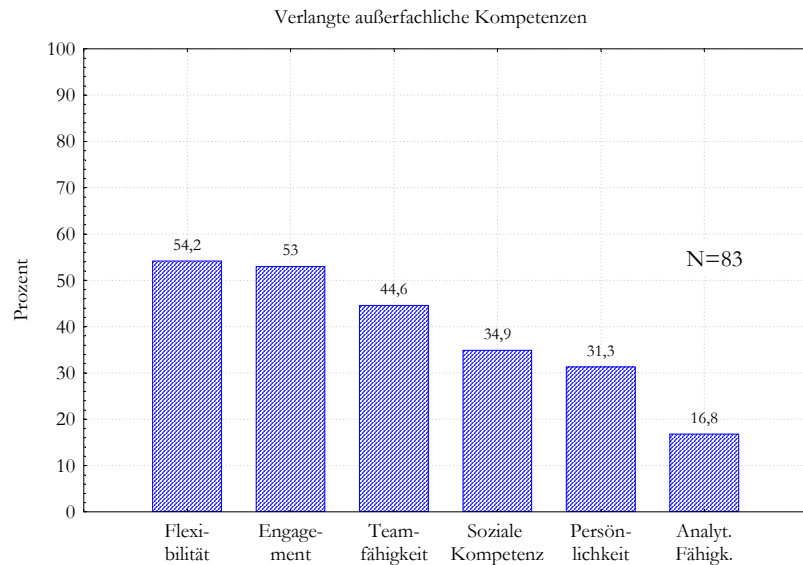


Abbildung 5.2: Anteil der Unternehmen, welche die jeweilige Kompetenz als Einstellungskriterium angaben (aus *Stellenreport 2000/2001*)

ist (Robertson & Makin, 1986). Zudem sollte ein Bezug zu dem intern erarbeiteten Anforderungsprofil bestehen, welches der Personalauswahl zugrunde liegt. Beides wird durch die gewählten Begriffe erschwert. Soweit man davon ausgehen kann, dass kommunizierte Anforderungen den tatsächlichen Erwartungen eines Unternehmens entsprechen, basieren diese auf einer wenig differenzierten Definition der relevanten Eigenschaften.

Man könnte dagegen halten, dass diese Stellenanzeigen eine eher politische Kommunikationsfunktion haben und nicht direkt von den verantwortlichen Personalfachleuten formuliert werden, sondern von der Abteilung für Öffentlichkeitsarbeit, welche die eigentlichen Inhalte unter anderen Gesichtspunkten modifizieren und ergänzen. Aus wissenschaftlicher Sicht erscheint die Darstellung außerfachlicher Kompetenzen für Führungsnachwuchs sehr allgemein und wenig konkret, hat aber wohl für die Praxis den Vorteil der Knappheit. Möglicherweise verbergen sich dahinter spezifischere Dimensionen. In der Praxis sind diese Benennungen allgemein verbreitet. Die gleichen Begriffe enthält zum Beispiel eine Befragung der Personalberatung Kienbaum bei Unternehmen, die sich für wichtige Eigenschaften von Führungskräfte interessieren (Schmidt-Klingenberg, 2002).

Wie bei den zuvor berichteten Untersuchungen war auch hier das Ge-

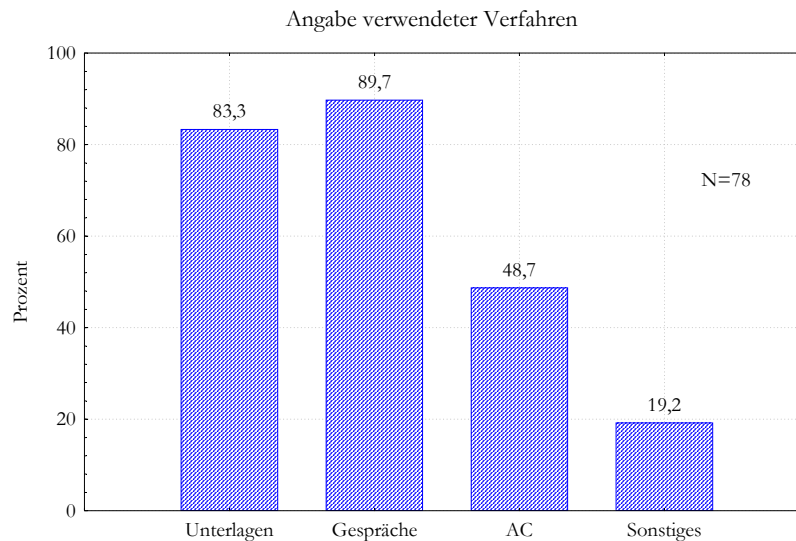


Abbildung 5.3: Anteil der Unternehmen, die das jeweilige Auswahlverfahren angeben (ausgewertet aus Angaben im *Stellenreport 2000/2001*).

spräch oder das Interview das am häufigsten genutzte Verfahren. Es wurde fast doppelt so häufig genannt wie das Assessment-Center. Seltener ist die Auswertung der Bewerbungsunterlagen seltener angegeben. Das liegt möglicherweise daran, dass dies ohnehin vorausgesetzt wird. Interessant wäre noch gewesen, aus welchen Bausteinen die Assessment-Center bestehen, aber das ließ sich auf dieser Grundlage nicht ermitteln.

5.3 Outcome-Evaluation im Konzern

Nachdem die Anforderungsmodelle an Führungsnachwuchs in Stellenanzeigen möglichst induktiv abgebildet wurden, soll nun ein Projekt geschildert werden, bei dem das theoriegeleitete Anforderungsmodell an Problemlöser in ein Kompetenzmodell für obere Führungskräfte eingearbeitet wurde. In einem mehrstufigen Prozess wurden neben dem wissenschaftlich begründeten Modell mehrere Perspektiven integriert. Ausgehend von dem Problemlösemodell wurde dieses erweitert um die besondere Unternehmenssituation, um die Erwartungen erfolgreicher Stelleninhaber sowie verschiedener Leiter der Personalabteilungen im Konzern. Diese Definition eines Anforderungsmodells für die anschließend geplante interne Personalentwicklung stellt eine

normative outcome evaluation dar, wobei qualitatives Vorgehen im Rahmen nominaler Gruppen eingesetzt wurde. Der Prozess und das Ergebnis werden im Folgenden beschrieben.

5.3.1 Hintergrund

Angeichts der Liberalisierung des Energiemarktes beschloss ein deutsches Energieversorgungsunternehmen eine Bestandsaufnahme und strategische Neuorientierung angesichts der eingetretenen und noch zu erwartenden Veränderungen, welche Wettbewerb ermöglichten und erforderten. Es wurde angenommen, dass diese zukünftig andere Kompetenzen von oberen Führungskräften verlangen als bisher. Deshalb sollte eine Potenzialanalyse den Status quo der derzeitigen Führungsebene in Bezug auf diese veränderten Anforderungen abbilden. Das sollte die fundierte Ermittlung des Qualifizierungsbedarfes und solide Planung von Entwicklungspfaden im Konzern ermöglichen. Daran sollten sich Führungskräfteentwicklungsprogramme anschließen. Dafür war ein operationalisierbares Kompetenzmodell für diese Führungskräfte nötig.

Die zuständige Konzernabteilung Führungskräfteentwicklung wollte dabei aufgrund früherer Erfahrungen wissenschaftliche Befunde neben praktischen Erfahrungswerten und internen Informationen berücksichtigen, weshalb in dieser Projektphase ein universitärer Berater engagiert wurde. Ein externes Beratungsunternehmen sollte später die Potenzialanalyse durchführen. Zu diesem Zweck wurden in einem mehrstufigen Prozess die verschiedenen Perspektiven gesammelt, Rückmeldungen eingeholt und mehrfach in verschiedenen Gruppen diskutiert, bis schließlich das endgültige Kompetenzmodell vorlag. Dazu wurde zunächst ein Ablaufplan sowie Zuständigkeiten der beiden Seiten festgelegt. Für jeden Schritt gab es schriftliche Entwürfe, ebenso schriftliche Kommentare, die sukzessive integriert wurden. Der Austausch erfolgte wegen der großen Entfernung über Email, Telefontermine und persönliche Treffen im Unternehmen.

Aus der Sicht des Unternehmens und laut vertraglicher Vereinbarung sollte die wissenschaftliche Beratung neben den genannten Zielen dabei unterstützen, dass die Potenzialanalyse gleichzeitig transparent, einfach und aussagekräftig gestaltet wird, um die Akzeptanz bei den Beteiligten zu optimieren. Zielsetzung ist, bei den oberen Führungskräften Potenzial für die nächsthöhere Führungsposition zu identifizieren. Diese Führungskräfte sind hinsichtlich ihrer Verantwortung, Führungsspanne, wirtschaftlichen Bedeutung und so weiter unterschiedlich. Damit stand Personalentwicklung im Mittelpunkt, weniger Personalauswahl. Die Potenzialanalyse soll dabei als handhabbares und flexibles Tool entwickelt werden. Am Ende werden die Teilnehmer

in ein Portfolio kategorisiert, wobei ein Abgleich der vorhandenen Kompetenzen und Potentiale mit den aus der Strategie abgeleiteten zukünftigen Anforderungen an Inhaber von Schlüsselfunktionen erfolgt.

5.3.2 Methode

Eine normative Outcome-Evaluation soll die Interessenträger dabei unterstützen, die Programmziele zu identifizieren, zu klären und zu bewerten. Sie hat politische und operative Funktionen. Auf politischer Seite geht es darum, Akzeptanz und Unterstützung der Maßnahme zu sichern, auf operativer Seite sollen Kriterien für die Programmentwicklung und -bewertung erarbeitet werden. Die wissenschaftliche Begleitung sollte in diesem Projekt unterstützen, mögliche Anforderungen an die Manager vorzuschlagen und einzugrenzen. Diese schließlich möglichst präzise abzubilden war das Ziel der Potenzialanalyse, wobei der Anforderungskatalog die Zielkriterien umfasst.

-
1. Bestimmung der Informationskategorien
 2. Strategische Situationsanalyse
 3. Nominale Gruppe I - Personalleiter
 4. Nominale Gruppe II - Geschäftsführer
 5. Verabschiedung im Vorstand
-

Tabelle 5.2: Ablaufschritte der Outcome-Evaluation zur Entwicklung eines Kompetenzmodells für Führungskräfte im Konzern.

Zur Bündelung und Priorisierung von Zielkriterien sind in der Programmevaluation verschiedene, unterschiedlich standardisierte quantitative und qualitative Techniken gängig, die den Entscheidungsprozess transparent machen und unterstützen. Im beschriebenen Projekt wurden dazu nominale Gruppen eingesetzt, die in mehreren Stufen das Ausgangsmodell unter verschiedenen Aspekten bewerteten und ergänzten. Der Ansprechpartner des Unternehmens und der wissenschaftliche Berater koordinierten gemeinsam den gesamten Prozess. Sie arbeiteten diese Informationen nach jeder Rückmeldephase durch die jeweilige nominale Gruppe ein, um mit dem überarbeiteten Modell in die nächste Stufe zu gehen. Den nominalen Gruppen waren zwei Phasen der Informationssammlung vorgeschaltet. Ausgangspunkt bildete das oben vorgestellte Anforderungsmodell, das um strategische Informationen aus Geschäftsberichten und internen Papieren angereichert wurde.

Dabei wurden zwei nominale Gruppen identifiziert und gebildet. Die erste Gruppe bestand aus acht Leitern der Personalabteilungen aus verschiedenen Unternehmensbereichen, die zweite Gruppe bildeten neun erfolgreiche Inhaber von Schlüsselpositionen im oberen Management, das heißt, Geschäftsführer in verschiedenen Geschäftsbereichen. Um die Meinung der Personalexperten zu sammeln und zu strukturieren, wurde eine Metaplansitzung mit acht Teilnehmern und zwei moderierenden Personen durchgeführt. Sie diente in erster Linie der Strukturierung, aber auch dem Einholen ergänzender Beiträge. Der Ablauf entsprach einer typischen Kurz Sitzung von drei Stunden. Im Mittelpunkt stand die Kartenabfrage, Sortieren, Ergänzung und Diskussion zum vorgestellten Kompetenzmodell.

Mit halbstandardisierten Einzelinterviews wurden die Beiträge der Geschäftsführer zum Ergebnis der Metaplansitzung eingeholt, indem sie zunächst das Modell schriftlich zugeschickt bekamen, zusammen mit den beiden Leitfragen, ob und wie weit sich die Inhalte ihrer Arbeit damit abdecken ließen und welche Ergänzung sie gegebenenfalls vorschlagen würden. Dies diente als Grundlage des persönlich oder telefonisch geführten Interviews. Dabei wurden insgesamt neun Personen befragt. Am Ende dieses Prozesses stand das Kompetenzmodell für obere Führungskräfte, das zuletzt im Vorstand präsentiert und verabschiedet wurde. In welchen Schritten dieses Kompetenzmodell entwickelt wurde, beschreiben die folgenden Abschnitte. Das Ergebnis für den Konzern wird unten kurz dargestellt.

5.3.3 Durchführung

Bis zur endgültigen Form des Kompetenzmodells umfasste der Prozess mehrere Stufen. In der ersten Stufe wurde das wissenschaftliche Kompetenzmodell präsentiert, mit betrieblichen Modellen verglichen, und für verschiedene Aspekte der Zusammenarbeit gemeinsam verbindliche Begriffsdefinitionen erarbeitet. Anschließend wurden strategische Informationen aus den Geschäftsberichten gesammelt, um insbesondere bereits erfolgten und zu erwarteten Veränderungen gerecht zu werden. Im dritten Schritt wurde dies zu einem vorläufigen Kompetenzmodell integriert, um es schließlich im vierten Schritt in der nominalen Gruppe zu strukturieren, Inhalte zu gewichten und zu verdichten. Die Inhaber von Schlüsselpositionen, hauptsächlich Geschäftsführer, kommentierten dieses Ergebnis, was im letzten Schritt in die endgültige Form einfluss, wie sie vom Vorstand verabschiedet wurde.

Im Sinne der verschiedenen Evaluationsperspektiven nach Chen (1990) sollte hier wie auch in den folgenden Schritten einerseits die Stakeholder-Perspektive berücksichtigt werden, um die Akzeptanz durch die Auftraggeber, Anwender und Betroffenen zu optimieren, andererseits wollte das Un-

ternehmen explizit die wissenschaftliche Perspektive einbeziehen, weil es sich dadurch eine konzeptuelle Anreicherung und bessere Fundierung des Vorgehens versprach. Aus diesem Grund war die wissenschaftliche Begleitung gewünscht, die Initiative für diese Zusammenarbeit ging vom Unternehmen aus.

1. Bestimmung der Informationskategorien Aufgrund der interdisziplinären Zusammenarbeit mussten zunächst gemeinsame Definitionen für einige Begriffe der weiteren Modellentwicklung erarbeitet werden, die sich sowohl auf die Eignungsdiagnostik allgemein sowie auf das Kompetenzmodell bezogen. Ausgangspunkt war die Potenzialanalyse. Diese wurde als ein Sammelbegriff für jedes Vorgehen verstanden, mit dem die berufsrelevanten Fähigkeiten eines Mitarbeiters eingeschätzt und zusammengefasst werden, üblicherweise ein System aus Fremd- und Selbsteinschätzung, strukturiertem Interview und Beratung. Weiterhin wurden Begriffe wie Potenzial, Kompetenz und Kompetenzbereiche diskutiert.

Besonders anspruchsvoll stellte sich für alle Beteiligten die Definition von Potenzial heraus. Es wurde als eine in die Zukunft projizierte Einschätzung der Kompetenzentwicklung gefasst und wies damit auf eine hohe Erwartungshaltung der Praktiker an die Diagnostik hin, die möglicherweise nur schwer einzulösen war. Im Unterschied dazu war Kompetenz die Summe der in der Vergangenheit erworbenen Fähigkeiten. Damit verbunden waren Zukunftskompetenzen, hier Kompetenzen, die zukünftig möglicherweise relevant sind. Deshalb konnte deren Relevanz aktuell noch nicht belegt werden, wie zum Beispiel internationale Kompetenz.

Daneben wurden konkrete Kompetenzbereiche umrissen, gesammelt und diskutiert. Ausgangspunkt waren zum einen das auf Seite 87 dargestellte eigene Kompetenzmodell als Angebot der wissenschaftlichen Perspektive, zum anderen die im Konzern bereits vorhandenen Kompetenzmodelle für Führungskräfte aus früheren Auswahl- und Entwicklungsmaßnahmen, sowie aktuell vorliegende Skizzen zu relevanten Kompetenzbereichen bei Führungskräften. Zunächst wurden die Definitionen für alle Kompetenzfelder abgeglichen. Häufig gab es verschiedene Überbegriffe mit ähnlichen Definitionen, aber auch ähnliche Begriffe, die auf praktischer und wissenschaftlicher Seite unterschiedlich gefasst waren.

Es gab viele Überschneidungen zwischen den Praktikermodellen und dem wissenschaftlichen Modell, insbesondere die interpersonellen Anforderungen, die auf Unternehmensseite als klassische Führungskompetenzen angesehen wurden, und strategische Aspekte, wie Zielausarbeitung und -Balancierung. Daneben sahen die Praktiker unternehmerische Kompetenzen und damit

Entrepreneurship als sehr wichtig an, deshalb floss noch ein wissenschaftliches Modell zum unternehmerischen Potenzial von Müller (1999) ein. Als interessant wurden die Bereiche der Informationsverarbeitung und des Selbstmanagements aus dem Problemlösemodell betrachtet. Zusätzlich wurden sehr viele fachliche und fachnahe Kompetenzen aus der Perspektive des Unternehmens genannt.

2. Strategische Situationsanalyse Das Kompetenzmodell sollte außerdem, wie eingangs erwähnt, den besonderen Rahmenbedingungen und strategischen Zielen des Unternehmens gerecht werden. Deshalb wurden Geschäftsberichte, interne Strategiepapiere sowie Status-quo-Analysen und Prognosen aus der Innen- und Außenperspektive gesammelt und strukturiert. Daraus sollten sich besondere Kompetenzschwerpunkte ableiten lassen. Die Strukturdimension waren nach Unternehmensnähe der relevanten Rahmenbedingungen aufgliedert, in interne Bedingungen, wie das Angebot, Ressourcen, Kostenfaktoren und so weiter, in marktbezogene Bedingungen wie Standort, Nachfrage, Mitbewerbern sowie Marktschwerpunkten und in äußere Rahmenbedingungen wie Politik, Technologie, Gesellschaft und Interessenverbände.

Es wurde angenommen, dass der günstige oder ungünstige Einfluss dieser Faktoren besonders von entsprechend qualifizierten und engagierten Mitarbeitern in den verschiedenen Funktionen und Ebenen des Unternehmens abhängt, also von den Inhabern sogenannter Schlüsselfunktionen oder Schlüsselpositionen. Daraus wiederum wurden die Anforderungen an diese Personen abgeleitet. Diese schließen fachliche Kenntnisse und Fertigkeiten ebenso ein wie persönliche Fähigkeiten von Führungskräften in Schlüsselfunktionen ein, da sie mit ihren persönlichen Voraussetzungen und Zielen entscheidenden Einfluss auf das Arbeitsumfeld und die Unternehmenskultur nehmen. Im Hinblick auf die Umsetzung der Unternehmensstrategie zeichnen sich Schlüsselfunktionen aus durch einen entscheidenden Einfluss auf die Gestaltung von Strukturen und Abläufen, auf das zielgerichtete Handeln der Mitarbeiter sowie die auf die Unternehmenskultur.

Damit war gleichzeitig eine Eingrenzung der Zielgruppe auf die Inhaber von Schlüsselfunktionen vorgenommen worden. Dabei wurden aus den Konzernzielen weitere relevante Kompetenzen für Führungskräfte abgeleitet, die nötig waren, um die Konzernstrategie umzusetzen. Insbesondere unkalkulierbare Marktentwicklungen können dazu führen, dass Konzernziele schwieriger zu erreichen sind. Deshalb muss das Kompetenzmodell auch für die Führungskräfte die Möglichkeit zur Modifikation und Erweiterung der Konzernstrategie beinhalten. Angesichts dieser postulierten stärkeren Eigenverantwortlichkeit und der Einflussmöglichkeiten dieser Personen wurde

ein weiteres Konzept eingeführt, die unternehmerische Kompetenz oder das sogenannte Entrepreneurship. Dies wurde in Anlehnung an Müller (1999) verstanden, welches motivationale und Persönlichkeitseigenschaften enthält, die besonders zum unternehmerischen Erfolg beitragen, wie zum Beispiel Verantwortungsbereitschaft und Risikotoleranz.

3. Nominale Gruppe I - Personalleiter Damit sammelte und bündelte die verantwortliche Personalabteilung und die wissenschaftliche Begleitung Informationen aus mehreren Quellen, nämlich aus dem wissenschaftlichen Modell, aus internen Entwürfen für ein Kompetenzmodell und aus den strategischen Analysen. Vor diesem Hintergrund entstand ein erster Vorschlag für ein Kompetenzmodell mit den vier Hauptdimensionen Problemlösekompetenz, Führung, Unternehmertum und Fachkompetenz. Dieser erste Vorschlag war wesentlich detaillierter und ausführlicher als für die endgültige Fassung geplant. Zum einen sollten durch detaillierte Definitionen Unklarheiten in den interdisziplinären Teams möglichst vorgebeugt werden, zum anderen diente dies als Vorlage für die erste nominale Gruppe, in der das Modell verdichtet und fokussiert werden sollte. Damit stand in diesem Schritt die Perspektive der Stakeholder im Vordergrund.

Dazu wurde die Metaplanmethode genutzt als qualitatives, strukturierendes Verfahren, mit dem die Teilnehmer ihre Beiträge auf Karten schreiben, diese gemeinsam diskutieren, sortieren, gewichten und ergänzen können. Wie üblich, moderierten zwei Personen die Metaplansitzung, je eine Vertreterin der Abteilung Führungskräfteentwicklung und der wissenschaftlichen Begleitung. Zur Vorbereitung bekamen alle Teilnehmer das vorläufige Kompetenzmodell zugeschickt, ebenso die Leitfragen, die auch die Gruppensitzung bestimmten. Zum einen sollte das Modell als Ganzes kommentiert werden, zum anderen Schwerpunkte gesetzt und einzelne Teilkompetenzen zu übergeordneten Dimensionen zugewiesen werden.

Die Teilnehmer waren acht Personalleiter aus verschiedenen, aber als bedeutsam angesehenen Geschäftsbereichen, deren Abteilungen in der Konzernzentrale untergebracht waren, wie auch die Abteilung Führungskräfteentwicklung. Alle hatten einen Hochschulabschluss, es handelte sich um vier Diplom-Kaufmänner, zwei Diplom-Psychologen, einen Juristen und eine Soziologin. Die Soziologin war die einzige Frau dieser Gruppe und führte als einzige einen Dokortitel. Die Metaplan-Sitzung dauerte drei Stunden und kann damit als Kurzsitzung angesehen werden.

Bei der Diskussion des Modells als Ganzes konnten die Teilnehmer eigene Kommentare und Ergänzungen formulieren und den vier Hauptdimensionen zuordnen. Im zweiten Schritt wurden Teilkompetenzen zu Teilkategorien

verdichtet und gewichtet. Eine wesentliche Ergänzung war die Teilkategorie Kreativität, die dem Unternehmertum zugeordnet wurde. Sie war im bisherigen Modell nicht berücksichtigt worden, und die Schwierigkeit der Operationalisierung wurde diskutiert. Dennoch sollte sie, auch unter normativen Gesichtspunkten, mit aufgenommen werden. Sehr viele Beiträge betrafen die Bezeichnung von Teilkategorien, die auch unter politischen Gesichtspunkten betrachtet wurden, da das Modell ja an verschiedene Interessengruppen kommuniziert werden und dort Akzeptanz finden musste. Dies spiegelt die Interdisziplinarität des Teams wieder. Daneben thematisierten einige Teilnehmer emotionale Intelligenz. Diese ließ sich als Erweiterung des Selbstmanagements bei der Problemlösekompetenz unterbringen.

Im Ergebnis wurden die vier Hauptdimensionen akzeptiert und beibehalten, mit leicht veränderten Bezeichnungen. Die Problemlösekompetenz und die Führungskompetenz enthielten je vier Teilkategorien, die unternehmerische Kompetenz und Fachkompetenz fünf Teilkategorien. Die Beiträge aus dieser ersten nominalen Gruppe wurden protokolliert und aufbereitet, um sie der nächsten nominalen Gruppe zur Diskussion zu stellen.

4. Nominale Gruppe II - Geschäftsführer Als zweite nominale Gruppe wurden die Inhaber von Schlüsselpositionen identifiziert, welche das Kompetenzmodell vor dem Hintergrund ihrer praktischen und persönlichen Erfahrungen beurteilen sollten. Als solche Schlüsselfunktionen wurden die Geschäftsführer der Konzernbereiche angesehen. Auch sie sollten zunächst allgemeine Kommentare angeben, daneben sollten sie beurteilen, wie weit das Modell derzeit und voraussichtlich in Zukunft berufliche Anforderungen widerspiegelt. Zu diesem Zweck bekamen sie zunächst das Modell zugeschickt, ebenfalls wieder mit Leitfragen, und es wurde ein persönlicher oder Telefontermin für ein halbstandardisiertes Interview vereinbart. Die jeweiligen Kommentare wurden eingearbeitet und beim nächsten Gespräch bereits präsentiert.

Insgesamt gab es keine grundlegenden Einwände gegen das Kompetenzmodell. Ähnlich wie bei den Personalleitern bezogen sich die Kommentare auf Details. Insbesondere die Bezeichnung der Problemlösekompetenz wurde als zu negativ angesehen, dies erinnere an Problemorientierung, wo Lösungsorientierung nötig sei. Deshalb wurde für diesen Kompetenzbereich die Bezeichnung Gestaltungskompetenz eingeführt, ebenso wurde die Fachkompetenz in bereichsspezifische Kompetenz umbenannt, um dies gegenüber Charakteristiken der Zuordnung zu Tarifgruppen zu unterscheiden, da es sich um Führungskräfte handelte. Das Endresultat dieser Runde wurde noch einmal an die Personalleiter zurückgemeldet.

I. Gestaltungskompetenz	III. Führungskompetenz
1. Informationsmanagement	1. Zielvereinbarung
2. Zielbildung	2. Mitarbeiterorientierung
3. Maßnahmenplanung	3. Kommunikation
4. Selbststeuerung	4. Personalentwicklung
II. Unternehmerische Kompetenz	IV. Spezialkompetenzen
1. Leistungsmotivation	1. Bereichsspezifische Kompetenz
2. Verantwortungsbereitschaft	2. Bereichsübergreifende Kompetenz
3. Risikobereitschaft	3. Veränderungskompetenz
4. Kreativität	4. Internationale Kompetenz
5. Durchsetzungsfähigkeit	5. Sonstige Kompetenzbereiche

Tabelle 5.3: Umsetzung eines Kompetenzmodells zur Potenzialeinschätzung oberer Führungskräfte im Konzern.

5. Verabschiedung im Vorstand Nachdem das Kompetenzmodell zur Potenzialeinschätzung bei oberen Führungskräften diese Schritte durchlaufen und seine endgültige Form erhalten hatte, wurde es vom Leiter der Konzernabteilung Führungskräfteentwicklung im Vorstand präsentiert und verabschiedet. Dieses Ergebnis des Kompetenzmodells mit seinen Haupt- und Teilkategorien wird im folgenden Abschnitt vorgestellt, eine Übersicht gibt Tabelle 5.3.

5.3.4 Ergebnisse

Das Kompetenzmodell zur Potenzialeinschätzung für obere Führungskräfte umfasst in seiner endgültigen Form, wie es im Konzern verabschiedet und publiziert wurde, vier Hauptdimensionen, die unterschiedliche Perspektiven der Beteiligten repräsentieren. Die Gestaltungskompetenz enthält vor allem die wissenschaftliche Perspektive. Die unternehmerische Kompetenz spiegelt ebenfalls eher wissenschaftliche Ansätze wider, allerdings vor dem Hintergrund der strategischen Situationsanalyse aus dem Unternehmen. Die Führungskompetenz als dritte Hauptdimension beinhaltet Aspekte des unternehmensinternen Kompetenzmodells sowie interpersoneller Fertigkeiten aus dem eigenen Anforderungsmodell. In der vierten Hauptdimension schließlich, den Spezialkompetenzen, sind besondere fachliche und nichtfachliche Anforderungen aus der Sicht des Unternehmens enthalten.

In der ersten Hauptdimension, der sogenannten *Gestaltungskompetenz* sind die vier Anforderungsdimensionen des Problemlösemodells von Dörner

(1986) untergebracht. Gestaltungskompetenz wurde hier als Fähigkeit definiert, in komplexen, dynamischen, neuen und vernetzten Situationen erfolgreich, das heißt zielführend zu agieren. Innerhalb der Gestaltungskompetenz unterscheidet man ähnlich wie bei Dörner Informationsgewinnung und -strukturierung, Zielausarbeitung und -balancierung, Maßnahmenplanung und Entscheidung sowie Selbststeuerung.

Die *unternehmerische Kompetenz* beinhaltet die eigenständige Ausrichtung jeder Führungskraft auf die Belange des Unternehmens, die durch persönliche und motivationale Eigenschaften geprägt ist. Sie enthält Leistungsmotivation als positive Einstellung zur Aufgabenbewältigung, Verantwortungsbereitschaft für die eigene Person, andere Aufgaben und andere Menschen. Risikobereitschaft als dritte Teilkategorie bestimmt den Umgang mit Entscheidungen unter Unsicherheit. Kreativität ist eine wichtige Voraussetzung, um im Unternehmen Prozesse zu gestalten. Durchsetzungsfähigkeit bezieht sich auf das erfolgreiche Durchsetzen eigener Aufgabeninteressen und umfasst kommunikative, kompetitive und kooperative Aspekte.

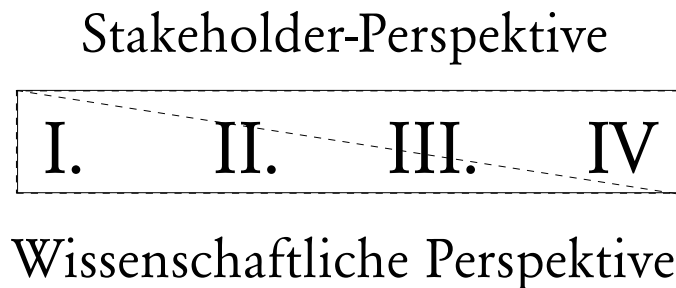


Abbildung 5.4: Wissenschaftliche und Anwenderperspektive in den vier Bereichen des Kompetenzmodells

Das dritte Kompetenzfeld *Führungskompetenz* beinhaltet die Mitarbeiterführung mit eher klassischen Führungsaufgaben und einem interpersonellen Schwerpunkt. Im Vordergrund steht die Zielerreichung mit anderen und das gemeinsame Gestalten. Die vier Teilkategorien sind Zielvereinbarung und -erreichung, Mitarbeiterorientierung, kommunikative Fähigkeiten sowie Personalentwicklung.

Schließlich wurden aus den besonderen Bedingungen und Anforderungen der Unternehmenssituation heraus als viertes die *Spezialkompetenzen* formuliert. Sie spielen zusätzlich zu den anderen Kompetenzen eine wichtige Rolle und setzen sich aus fünf Teilkategorien zusammen. Dabei handelt es sich um die bereichsspezifische Kompetenz, bereichsübergreifende Kompetenz, Ver-

änderungskompetenz, internationale Kompetenz sowie sonstige Kompetenzbereiche wie EDV-Kenntnisse oder Projektmanagement.

5.4 Metaanalyse: Persönlichkeit

Die Metaanalyse wird häufig als Evaluationsstrategie eingesetzt. Ebenso wie alle anderen Methoden kann auch sie je nach Fragestellung in verschiedenen Bereichen verwendet werden. Ihr Schwerpunkt liegt generell eher der summativen Evaluation, da die Ergebnisse einer Metaanalyse eher für eine integrative Zusammenschau der Ergebnislage genutzt werden, als für eine Rückkopplung in eine Maßnahme. In der Eignungsdiagnostik werden Metaanalysen vor allem in der Impact-Evaluation eingesetzt, wenn es darum geht, die Kriteriumsvalidität von diagnostischen Verfahren und Konstrukten festzustellen. Gleichzeitig werden die Ergebnisse solcher Impact-Evaluationen in der Praxis genutzt, um Kompetenzmodelle festzulegen, die erfasst werden sollen, so dass diese Ergebnisse außerdem einen Beitrag zur Outcome-Evaluation liefern.

Bisher gibt es im Vergleich etwa zur Personaleignungsdiagnostik recht wenige Metaanalysen, die sich direkt auf Befunde aus der Führungsforschung beziehen. Ebenso wie in anderen Bereichen werden diese zunehmend kritisiert. Die Metaanalysen erscheinen nach einer systematischen Durchsicht bezüglich ihrer Methoden und Ergebnisse als genauso inkonsistent wie die Einzelstudien, und können die widersprüchliche Befundlage nicht erklären (Klenke, 1994). Dennoch sollen hier einige Studien integriert werden, die sich damit beschäftigen, welche Persönlichkeitsmerkmale mit Führungserfolg zusammenhängen, um auch einen Rückschluss auf Kompetenzmodelle ziehen zu können. Der Ablauf dieser Metaanalyse wird im Folgenden vorgestellt.

5.4.1 Vorbereitung

Typischerweise ist bei einer Metaanalyse zunächst die Forschungsfrage festzulegen. Hier soll geklärt werden, welchen Beitrag die Persönlichkeitsmerkmale aus der Gruppe der Big Five (McCrae & Costa, 1987) zur Vorhersage von Führungserfolg leisten, da sich derzeit die Persönlichkeitsdiagnostik überwiegend an diesem Modell orientiert. Damit ist bereits der Prädiktor weitgehend festgelegt, der in der Ausprägung auf einer oder mehrerer dieser Persönlichkeitsdimensionen besteht, typischerweise operationalisiert in den entsprechenden Skalen oder vergleichbaren Maßen. Das Berufserfolgskriterium besteht in Maßen für Führungserfolg, deren Operationalisierung uneinheitlicher ist. Inzwischen wird Führungsleistung zunehmend häufiger anhand der

Autoren	Jahr	Stichprobe
Boudreau, Boswell & Judge	2001	1885 US- Führungskräfte 1871 europ. Führungskräfte
Caligiuri	2000	143 US-Führungskräfte
Church & Waclawski	1998	235 US-Offiziere
Conway	2000	1567 US-Führungskräfte
de Jong, Bouhuys & Barnhoorn	1999	58 internat. Führungskräfte
Gellatly & Irving	2001	79 US- Führungskräfte
Judge, & Bono	2000	174 US-Führungskräfte
Kickul & Neuman	2000	67 Freiwillige
McCredie & Shackleton	2000	70 UK-Führungskräfte 235 UK-Führungskräfte
Motowidlo, Burnett et al.	1996	60 US-Führungskräfte
Orpen	1996	136 Austr. Führungskräfte
Ployhart, Lim & Chan	2001	1259 Asiat. Nachwuchskräfte
Robertson, Baron et al.	2000	437 UK-Führungskräfte
Stricker & Rock	1998	223 US-Offiziere
Taggar, Hackett & Saha	1999	480 US-Freiwillige
Thomas, Dickson & Bliese	2001	818 US-Offiziere

Tabelle 5.4: Für die Metaanalyse der Kriteriumsvalidität von Big-Five-Dimensionen verwendete Arbeiten.

in letzter Zeit entwickelten Führungsmodelle (zum Beispiel Conway, 2000) operationalisiert.

Der nächste Schritt beschreibt die Suchstrategie. Als Suchraum diente hauptsächlich die von der APA unterhaltene Literaturdatenbank *PsycInfo*, die inzwischen auch die meisten deutschen Publikationen führt. Verschiedene Suchbegriffe fragten in verschiedenen Kombinationen Titel und Abstracts ab: *leadership*, *management*, *managerial*, *executive* für die Zielgruppe, *personality*, *trait* für die Prädiktoren und *performance*, *success*, *criteria*, *predict*, *validity* für die Vorhersage von Berufserfolg. Diese Suchbegriffe waren sehr allgemein, da unbrauchbare Treffer ein kleineres Problem darstellten als wichtige Studien wegen zu spezieller Suchbegriffe zu übersehen, weil die gesamte Treffermenge überschaubar war. Außerdem sollten alle Studien nicht früher als im Jahr 1996 publiziert sein, um eine möglichst aktuelle Befundlage abzubilden. Damit ist ein Zeitraum von großzügig bemessenen fünf Jahren umfasst, was in vielen Reviews üblich ist.

Anhand der Abstracts ließ sich schnell feststellen, ob eine Studie tatsächlich Informationen im Sinne der Fragestellung enthielt, und sie wurde angefordert. Davon wurden alle Studien einbezogen, die bis zu einem bestimmten

Zeitraum vorlagen. Einige Treffer waren redundant, wie zum Beispiel Promotionen, die ebenfalls in Zeitschriften veröffentlicht waren. Unbedingt sollte die Studie mindestens einen Kennwert zum Zusammenhang zwischen einer Big-Five-Persönlichkeitsdimension und einem Führungserfolgsmaß enthalten, außerdem Untersuchungsdesign und Stichprobe beschreiben.

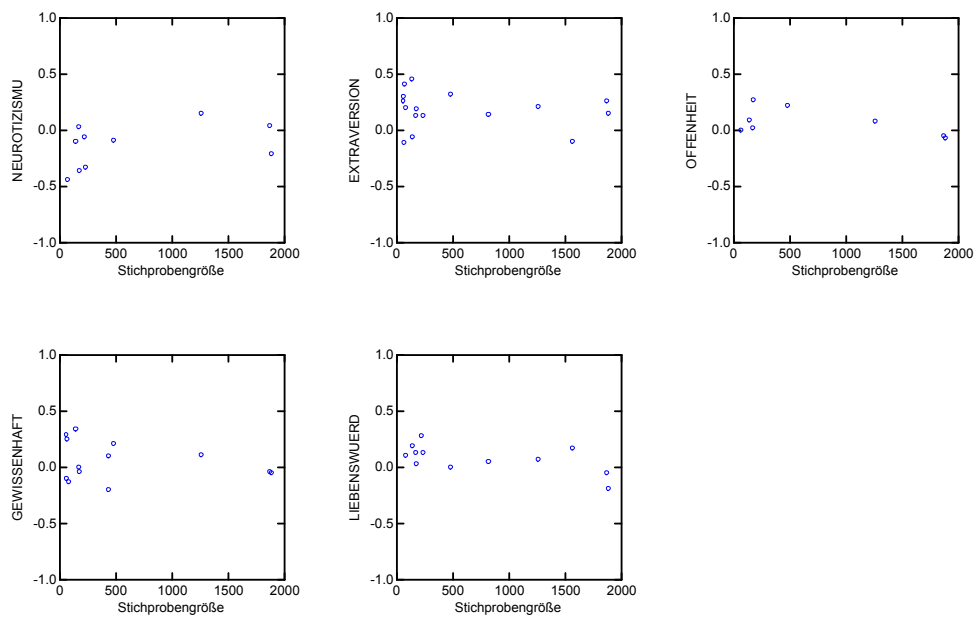


Abbildung 5.5: Streuung der Korrelationen mit Berufserfolg für jede der fünf Persönlichkeitsdimensionen. N Gesamt = 10398.

Nach diesen Vorgaben ließen sich 16 Studien zusammenstellen, manche davon enthielten mehrere Stichproben, etwa bei einem Vergleich europäischer und amerikanischer Führungskräfte. In diesem Fall wurden die Kennwerte für beide Stichproben getrennt aufgenommen. Die meisten Erhebungen fanden in Realsituationen statt, meist im Rahmen von Personalentwicklungsmaßnahmen. Schließlich wurden die Daten in eine Statistikdatei eingegeben, in der verschiedene Informationen festgehalten sind. Dies umfasst die Literaturangabe der Studie, Stichprobengröße, Operationalisierung von Prädiktoren und Kriterien, sowie die Korrelationskoeffizienten. Es wurden die unkorrigierten Korrelationskoeffizienten eingetragen, sofern diese berichtet waren. Eine kurze Übersicht über die Studien befindet sich in Tabelle 5.4, die genauen Angaben sind im Literaturverzeichnis separat aufgeführt.

	Neurot.	Extrav.	Offenh.	Gewissenh.	Verträglichk.
Gew. mittl. r	-.12	.15	.00	.00	.00
Ung. mittl. r	-.17	.18	.01	.01	.01
Anz. Studien	10	16	8	13	11
Total N	6504	9071	6048	7119	8729
CI .95 Mean	$-.05 > r > -.29$	$.27 > r > .09$	$.17 > r > -.03$	$.16 > r > -.05$	$.15 > r > -.01$

Tabelle 5.5: Übersicht über die Ergebnisse der Metaanalyse. Die mittleren Korrelationen basieren auf Fishers-Z-transformierten Werten.

5.4.2 Auswertung

Nachdem die Kennwerte, in diesem Fall Produkt-Moment-Korrelationen, erfasst wurden, müssen die Daten für weitere Auswertungen aufbereitet werden. Dazu sind mehrere Entscheidungen zu treffen, die in der Literatur zur Metaanalyse von eignungsdiagnostischen Untersuchungen umstritten sind (Hermelin & Robertson, 2001). Zunächst ist festzulegen, ob und wie die Korrelationen um eingeschränkte Varianz des Kriteriums und des Prädiktors korrigiert werden müssen, dazu muss eine Schätzung der Varianz für die Population bekannt sein. Außerdem ist zu entscheiden, ob die Werte um ihre Reliabilität minderungskorrigiert werden.

Beides führt meist dazu, dass die neuen Werte, mit denen weitergerechnet wird, wesentlich höher sind als die den Studien entnommenen. Dazu muss aber die Varianz beziehungsweise Einschränkung der Varianz in der Untersuchungstichprobe bekannt sein, was nicht alle Studien berichten. Dasselbe gilt für die Minderungskorrektur anhand der eingeschränkten Reliabilität. In einigen Studien wurde bereits die minderungskorrigierte Korrelation berichtet. Trotz der Möglichkeiten sprechen sich sowohl Schmidt & Hunter (1998) wie auch Bobko, Roth & Potowsky (1999) dafür aus, die Korrelationskoeffizienten nicht zu korrigieren, sondern mit den unkorrigierten Werten zu arbeiten, die auch dem „operational use“ eher entsprechen. Die Korrelationskoeffizienten wurden mit Vorzeichen ausgewertet, das heißt, dass hohe negative und positive Zusammenhänge bei einem Persönlichkeitsfaktor sich zu einem niedrigen Gesamtzusammenhang kompensierten. Manche Arbeiten aggregieren hier einfach den Betrag der Koeffizienten, doch das erscheint hier nicht sinnvoll. Jedoch wurde das Vorzeichen für Korrelationen mit *emotionaler Stabilität* gewechselt, wenn dies explizit so angegeben war, um es den restlichen Korrelationen mit *Neurotizismus* in der inhaltlichen Ausrichtung anzugleichen.

Da Pearson Produkt-Moment-Korrelationen nicht intervallskaliert sind, müssen sie vor weiteren Auswertungen in Fishers Z-Werte transformiert werden. Für diese kann das arithmetische Mittel, gewichtet nach Stichprobengrößen oder ungewichtet bestimmt werden, das dann wieder rücktransformiert wird, sowie das zugehörige Konfidenzintervall, das den Wert 0 nicht abdecken sollte. Ob das einen entscheidenden Unterschied ausmacht, wird diskutiert. Nach Hunter & Schmidt (1990) überschätzt dieses Vorgehen Korrelationen leicht, nach Fricke & Treinies (1985) macht es keinen Unterschied, während Rosenthal & DiMatteo (2001) dazu raten. Jedoch ist es methodisch korrekt, vor weiteren Berechnungen diese Transformation vorzunehmen, und das Resultat wieder zurück zu transformieren, deshalb soll es auch hier praktiziert werden.

5.4.3 Ergebnisse

Es wurden 16 Studien mit insgesamt 18 Untersuchungsstichproben ausgewertet. Für jede Big-Five-Dimension lagen maximal 16 (Extraversion) und mindestens 8 (Offenheit) Korrelationskennwerte vor. Die Gesamtstichprobe umfasste 10398 Teilnehmer. In ihrem Überblicksartikel zum Stand der Metaanalyse empfehlen Rosenthal & DiMatteo (2001) für Korrelationskoeffizienten zunächst eine grafische Aufbereitung der Ergebnisse, zur besseren Übersicht, und um Ausreißer besser abschätzen zu können. Studien mit auffälligen Werten sollte man noch einmal durchgehen, um möglicherweise besondere Merkmale zu identifizieren. Ein praktischer Vorschlag besteht darin, die Kennwerte gegen die Stichprobengröße abzutragen und als Scatterplot ausgeben zu lassen. Abbildung 5.5 zeigt diese Streuung der Korrelationen mit Berufserfolg für jede der fünf Persönlichkeitsdimensionen. Diese zeigt keinen sichtbaren Zusammenhang zwischen der Höhe der Korrelationen und der Stichprobengröße.

Weiterhin wurden die Korrelationskoeffizienten für jede Persönlichkeitsdimension nach Fisher Z-transformiert, und die gewichteten und ungewichteten arithmetischen Mittel der Z-Werte gebildet, anschließend erfolgte wieder ein Rücktransformation. Diese mittleren Effektgrößen, hier Korrelationen, sind meist das wesentliche Ziel einer Metaanalyse, können jedoch nur in Bezug zur Varianz der Kennwerte interpretiert werden. Diese wurde in Form des Konfidenzintervalls der Mittelwerte abgebildet. Diese Ergebnisse sowie der Umfang der Stichproben für jede Persönlichkeitsdimension befinden sich in Tabelle 5.5. Die Mittelwerte, die auf einer nach Stichprobengrößen gewichteten Berechnung beruhen, liegen im Betrag leicht unter den ungewichteten Mittelwerten.

Betrachtet man die Mittelwerte allein, gibt es für den Zusammenhang

von Neurotizismus und Extraversion mittlere Effekte nach der Konvention von Cohen (1988). Es ist also eine hohe Extraversion sowie eine hohe emotionale Stabilität günstig für Führungserfolg, während die anderen drei Persönlichkeitsdimensionen Offenheit für neue Erfahrungen, Gewissenhaftigkeit und Liebenswürdigkeit keine Rolle spielen. Auch vor dem Hintergrund des geschätzten Konfidenzintervalls für eine Wahrscheinlichkeit von 95 Prozent betrachtet, schließt dieses für die ersten beiden Dimensionen Null-Korrelationen nicht ein, während dies für die restlichen Persönlichkeitsmerkmale der Fall ist. Für die Extraversion umfasst sogar das gesamte Konfidenzintervall noch mittlere Effektstärken, die nach der Einteilung von Cohen im Bereich $.10 > r > .30$ liegen.

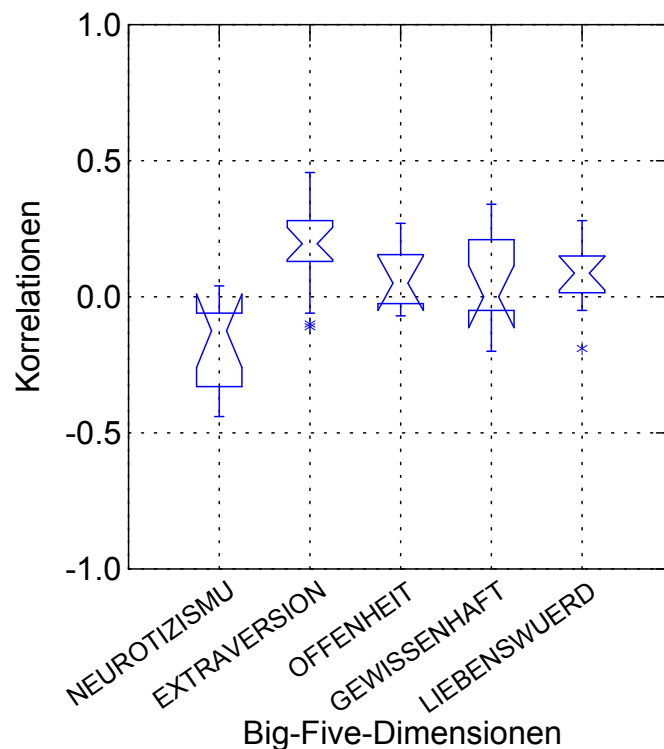


Abbildung 5.6: Konfidenzintervall der Mediane der Korrelationen von Persönlichkeitsfaktoren mit Berufserfolg für jede der fünf Persönlichkeitsdimensionen.

Mit den Ergebnissen für Extraversion und Neurotizismus entsprechen diese Ergebnisse den Befunden für den Zusammenhang mit Berufserfolg bei den meisten Berufsgruppen, während Gewissenhaftigkeit im Unterschied dazu bei Führungskräften eine kleinere Rolle spielt. Nach Robertson, Baron, Gibbons,

MacIver und Nyfield (2000) korreliert Gewissenhaftigkeit für Erfolgskriterien, welche die direkte Arbeitsleistung betreffen, kaum mit Führungserfolg, während er einen negativen Zusammenhang mit Karriereindikatoren fand. In der eigenen Metaanalyse wurden unterschiedliche Berufserfolgskriterien aggregiert, mit Schwerpunkt auf Einschätzungen der Arbeitsleistungen. Damit hat das Kriterium eine größere Bandbreite und ein höheres Aggregationsniveau, ist jedoch gleichzeitig weniger präzise. Bei einer größeren Stichprobe wäre es interessant, die Auswertung nach unterschiedlichen Erfolgskriterien aufzuteilen sowie nach Stichprobeneigenschaften, die aus unterschiedlichen Branchen und Führungsebenen stammen.

5.5 Fazit

In diesem Kapitel wurden drei eigene Studien vorgestellt, die aus unterschiedlichen Perspektiven Beiträge zu einer Outcome-Evaluation in der Management-Diagnostik liefern können, die ja darin besteht, Zielkriterien zu definieren. Eine Gruppe von Zielkriterien sind die Eigenschaften, die mit der Managementdiagnostik abgebildet werden sollen.

Eine Auswertung von Stellenanzeigen bei 83 Unternehmen bildete fachliche und nichtfachliche Anforderungsprofile aus der Perspektive der Praxis ab, wie sie gegenüber Bewerbern kommuniziert werden. Diese erscheinen relativ allgemein, und weisen nur wenig inhaltliche Bezüge zu psychologischen Eigenschaftskonstrukten auf. Dennoch bringen sie implizite Eignungstheorien zum Ausdruck und sind von hoher Relevanz, weil diese die Personalarbeit prägen.

So befragt die Personalberatung Kienbaum regelmäßig Unternehmen nach Persönlichkeitseigenschaften, die ihrer Ansicht nach für angehende Führungskräfte wichtig sind (Schmidt-Klingenberg, 2002). Dort wurden alle in der Auswertung der Stellenanzeigen gefundenen Fähigkeiten aufgeführt. Daneben spielen dort für die Unternehmen weitere Eigenschaften eine wichtige Rolle: Lernbereitschaft, Zielorientierung, Belastbarkeit, Selbstkritik, Konfliktfähigkeit, Führungspotenzial, Durchsetzungsvermögen und Risikobereitschaft. Außer der Belastbarkeit, die in der Metaanalyse von Persönlichkeitskonstrukten aufgeführt wurde, und dem tautologischen Begriff Führungspotenzial finden sich die restlichen genannten Eigenschaften in dem Kategoriensystem wieder, das vor dem Hintergrund des Problemlösemodells von Dörner entwickelt wurde in Verbindung mit interpersonellen Fertigkeiten (vergleiche Seite 87).

Die Dokumentation einer unternehmensinternen Entwicklung eines Kompetenzmodells zeigt eine Möglichkeit, die Praxisperspektive mit einer wis-

senschaftlichen Perspektive zu verbinden und betont die Notwendigkeit einer kontinuierlichen Abstimmung zwischen Evaluatoren und Auftraggebern oder Stakeholdern, damit die Ergebnisse für alle Beteiligten weiterverwendet werden können. Mit der besonderen Betonung der Problemlösekompetenz hebt sich dieses Modell zunächst von den vorher berichteten Praxismodellen ab. Doch die Unternehmensberatung Boston Consulting Group führte eine Interviewstudie mit rund 80 Experten und Führungskräften durch, wonach „strategisches, vernetztes Denken“ an der Spitze der Erfolgsfaktoren steht (Schmidt-Klingenberg, 2002).

Die Metaanalyse schließlich bildet die Bedeutsamkeit wissenschaftlicher Persönlichkeitskonstrukte als Bestandteil von Anforderungsprofilen für den Berufserfolg bei verschiedenen Stichproben von Führungskräften ab und gibt erste Hinweise darauf, welche davon für weitere diagnostische Maßnahmen wertvoll sein könnten. Ergänzend wäre hier noch eine Metaanalyse für kognitive Leistungsmerkmale als Prädiktoren sinnvoll, in dieser Arbeit wird dies bereits in einer eigenen Originalstudie bearbeitet.

Kapitel 6

Treatment-Evaluation

6.1 Übersicht

Im vorherigen Kapitel wurde ein Beispiel umgesetzt, wie man aus wissenschaftlicher Perspektive für die Managementdiagnostik zu einem Anforderungsmodell gelangen könnte, das sich auch als Grundlage einer normativen Outcome-Evaluation nutzen ließe. Kontrastierend dazu enthält dieses Kapitel eine Evaluationsstudie in der Praxis, in der die Stakeholder-Perspektive bei einer normativen Treatment-Evaluation im Vordergrund stand. Die Auftraggeber, die Personalabteilung eines Technologie-Konzerns, wünschten post hoc eine Evaluationsstudie ihrer internen Managementdiagnostik bei Führungskräften, von der Teilnehmerdaten vorlagen.

Es war Aufgabe der wissenschaftlichen Begleitung, zunächst in einer ersten, qualitativen Phase die Entwicklung von Evaluationskriterien durch die Auftraggeber zu unterstützen, um in einer zweiten, quantitativen Phase statistische Auswertungsstrategien vorzuschlagen und umzusetzen. Das diagnostische Instrumentarium beinhaltete unter anderem ein computergestütztes Verfahren zur Erfassung der Problemlösefähigkeit und sollte insgesamt fünf Anforderungsdimensionen abdecken, von denen drei problemlöserrelevant waren. Die anderen beiden Verfahren bezogen sich auf soziale Kompetenz und Persönlichkeit.

Neben den Auswertungsstrategien im Sinne einer normativen Treatment-Evaluation wurden die Daten für eine kausative Impact-Evaluation im Sinne des wissenschaftlichen und des betrieblichen Anforderungsmodells modelliert. Im Mittelpunkt stand die Modellierung und Begleitung der Urteilsprozesse auf verschiedenen Ebenen. Da der Datensatz die Voraussetzungen dafür erfüllte, ließen sich diese Varianten teilweise mit multivariaten statistischen Verfahren abbilden und prüfen. Weitere Zielkriterien wurden ebenfalls sta-

tistisch modelliert und geprüft. Die gesamte Auswertung musste den besonderen Anforderungen der Praxissituation gerecht werden, die unmittelbar relevante Informationen aus der Anwendung liefert und gleichzeitig durch die Erwartungen der Auftraggeber und methodische Einschränkungen geprägt ist.

6.2 Hintergrund

Im Sinne einer Diversifikationsstrategie wollte das Unternehmen neue Geschäftsfelder erschließen und längerfristig dort Schwerpunkte setzen. Die betrieblichen Ressourcen erhielt es teils durch die Übernahme entsprechender Unternehmen, teils wurden diese neuen Geschäftsbereiche selbst aufgebaut. Dafür wollte das Unternehmen neue Positionen mit entsprechender Führungsverantwortung besetzen, wobei es noch in der Planungsphase zunächst unternehmensintern nach geeigneten Anwärtern suchen wollte, die längerfristig entsprechende Verantwortung übernehmen könnten. Deshalb wurden in zwei aufeinanderfolgenden Jahren mit vorgeschlagenen Führungskräften des Konzerns einige eignungsdiagnostische Verfahren durchgeführt, aus deren Ergebnisse außerdem allgemeine Entwicklungsempfehlungen für die weitere Karriereplanung der Teilnehmer resultieren sollte.

Die zentrale Personalabteilung beauftragte zu diesem Zweck mehrere Personalberatungsunternehmen, welche die eignungsdiagnostischen Verfahren durchführen sollten. Jede Personalberatung war für ein oder mehrere Verfahren verantwortlich, die immer alle Kompetenzbereiche abdecken sollten. Gemeinsam mit diesen Beratungen wurde zunächst ein Anforderungsmodell erarbeitet, das vor allem außerfachliche Kompetenzen formulierte. Anschließend kamen aus dem Unternehmen Vorschläge für Teilnehmer an dem Auswahlverfahren. Die Durchführung oblag den Beratungsunternehmen, koordiniert wurde dies von der Personalabteilung des Unternehmens. Die Integration der Daten erfolgte wiederum bei einer der beteiligten Beratungen.

6.2.1 Praxisziele

Nachdem die Daten aus den eignungsdiagnostischen Verfahren im ersten Jahr angefallen waren, wünschte der Konzern auch eine Evaluation des Verfahrens anhand der vorliegenden Daten. Gemeinsam mit der Personalabteilung mussten deshalb im Rahmen der wissenschaftlichen Begleitung evaluierbare Ziele und Erwartungen an die Maßnahme identifiziert werden, um im nächsten Schritt Strategien für ihre Überprüfung zu entwickeln. Zunächst hatten die Auftraggeber sehr grundlegende Wünsche, wie eine integrative Doku-

mentation des Ablaufs und eine deskriptive Darstellung der Ergebnisse in der Gesamtheit und aufgegliedert nach verschiedenen Aspekten wie beteiligte Geschäftsbereiche, Resultate aus den beiden Jahren und so weiter. Diese Auswertungen und Darstellungen erfüllen die Dokumentationsfunktion von Evaluationen und sollen trotz ihres Umfangs hier nicht weiter ausgeführt werden.

Darüber hinaus wünschte sich das Unternehmen aussagekräftigere Resultate darüber, wie gut denn die Maßnahme gewesen sei, zum Beispiel vor dem Hintergrund des Anforderungsmodells. Daneben sollten Informationen über Zusammenhänge zwischen Anforderungsdimensionen und Verfahren zur Verfügung gestellt werden und eventuelle Urteilertendenzen geprüft werden. Ein weiteres Ziel war es, den gesamten Ablauf des mehrfach gestuften diagnostischen Prozesses zu modellieren und Vorschläge für eine systematische Urteilsbildung zu machen. Diese sehr allgemeinen Vorgaben wurden in einer qualitativen Phase spezifiziert und abgestimmt, bis schließlich ein Bündel von Evaluationszielen vorlag, das für den Auftraggeber relevant und für die wissenschaftliche Begleitung operationalisierbar war.

6.2.2 Betriebliches Anforderungsmodell

Dem gesamten diagnostischen Instrumentarium lag ein Anforderungsmodell zugrunde, das wesentliche außerfachliche Kompetenzen benannte, die das Unternehmen bei zukünftigen Geschäftsführern erwartete und die durch die diagnostischen Verfahren erfasst werden sollten. Diese waren lose an die *Big Five*-Persönlichkeitsfaktoren (McCrae & Costa, 1987) angelehnt, und vor dem Hintergrund der strategischen Unternehmensziele und den besonderen Anforderungen der neuen Geschäftsbereiche modifiziert worden. Vertreter der Personalabteilung und der Beratungsunternehmen hatten sie in mehreren Metaplanrunden formuliert.

1. Mit **(Führungs-)Persönlichkeit** ist ein unmittelbarer positiver Eindruck und eine positive Ausstrahlung gemeint, die sich in entsprechenden Denkmustern und Wertvorstellungen niederschlägt, die vor allem mit Selbstvertrauen und Autorität als wesentlichen Eigenschaften verbunden sind.
2. **Selbstmanagement** als zweite Dimension umfasst individuelle, aufgabenbezogene Eigenschaften wie Effektivität und Effizienz, ein hohes Ausmaß an Zielorientierung, Disziplin sowie Planungskompetenz. Im Vordergrund stehen dabei Sach- und Ergebnisorientierung.
3. Die Dimension **Systemisches Denken** ist am ehesten mit der Fähigkeit zum komplexen Problemlösen zu vergleichen, dies beinhaltet hier

ganzheitliches Denken in vernetzten, sich weiterentwickelnden Systemen, sowie die Fähigkeit, sich unter Berücksichtigung aller externen und internen Faktoren Veränderungen vorzustellen.

4. **Soziale Kompetenzen** als vierte Anforderung verlangt eine konstruktive Zusammenarbeit mit Menschen unter Einsatz adäquater Führungsphilosophie, die Ausrichtung von Menschen auf Strategien und Ziele sowie die Einflussnahme, und Motivation aller Menschen im betrieblichen Umfeld.
5. Zuletzt beinhaltet **unternehmerisches Handeln** die Fokussierung auf unternehmensrelevante Schwerpunkte, Entscheidungen und deren Durchsetzung vor dem Hintergrund der Unternehmensziele sowie die Fähigkeit, alle Aktivitäten auf den unternehmerischen Erfolg auszurichten.

6.3 Durchführung

Vor der eigentlichen Auswertung im Rahmen der Evaluationsstudie wird hier kurz vorgestellt, welche Verfahren in der Managementdiagnostik eingesetzt wurden, welche Daten daraus anfielen, in welcher Form sie vorlagen und welche Informationen neben den diagnostischen Ergebnissen zur Verfügung standen. Auch der mehrfach gestufte Ablauf ist kurz beschrieben. Schließlich wird erläutert, welche Evaluationsstrategie nach Chen gewählt, welche Teilfragen abgeleitet und wie sie statistisch operationalisiert wurden.

6.3.1 Verfahren

Im ersten Jahr durchliefen alle Teilnehmer der ersten Auswahlstufe drei Verfahren, Module eines computergestützten Postkorbs, als Persönlichkeitstest den Myers-Briggs Typenindikator MBTI (Briggs & Briggs-Myers, 1991), sowie einen internen halbstandardisierten Fragebogen zur Selbsteinschätzung beruflicher Interessen und Motive. Die Teilnehmer, die in die zweite Auswahlstufe kamen, absolvierten ein Assessment-Center. Für die Teilnehmer im zweiten Jahr kam in der ersten Auswahlstufe ein Vorgesetztenrating hinzu.

Die *Postkorb-Übung* ist ein typischer Baustein im Assessment-Center und gilt mit ihrer Struktur als repräsentativ für Managementaufgaben. Üblicherweise wird sie als Einzelübung durchgeführt. Als Material müssen Schriftstücke bearbeitet werden, indem der Proband geeignete Maßnahmen einleitet, Entscheidungen trifft und Termine koordiniert. Ergänzt wird die Instruktion durch weitere Zusatzinformationen während der Bearbeitung. Der

computergestützte Postkorb erweitert diese Übung um einen computergestützten Ablauf und eine automatische Auswertung. Er wurde als Windows-Programm mit der typischen Benutzeroberfläche umgesetzt. Das hier eingesetzten Verfahren wertet verschiedene Aspekte aus, zunächst die Bearbeitungsleistung in der verfügbaren Zeit, außerdem Prioritätensetzung, Delegation, Terminkoordination und Erkennen von Zusammenhängen im Abgleich zur Ideallösung. Neben dieser Zielerreichung wurde auch die Bearbeitungsstrategie betrachtet.

Generell ist der Postkorb weit verbreitet und wird von Teilnehmern gut akzeptiert. Leider liegen im deutschen Sprachraum nur wenige Untersuchungen vor. Ein wesentlicher Schwachpunkt vieler in der Praxis eingesetzter Postkorbübungen ist eine unzureichende Standardisierung (Musch & Lieberei, 1997). Häufig fehlt ein Bezugssystem und die Auswertung beruht auf freier Eindrucksbildung. Dadurch ist das Verfahren anfällig gegenüber Urteilstendenzen.

Der *Myers-Briggs Typenindikator* basiert auf C.G. Jungs Überlegungen zur Wahrnehmung als psychologischem Erkenntnisvorgang und zur Beurteilung als Gesamtheit aller Prozesse, in denen Wahrnehmungen zu Schlussfolgerungen werden. Unterschiedliche Wahrnehmungen und Beurteilungen führen zu Unterschieden in Reaktionen, Wertvorstellungen, Motivationen, Fähigkeiten und Interessen, wobei sich nach Präferenzen vier verschiedene Typen unterscheiden lassen, die vier bipolaren Testskalen entsprechen. Die erste kennzeichnet Extraversion versus Introversion, die zweite sinnliches versus intuitives Wahrnehmen, in der dritten Skala ist analytisches versus gefühlsmäßiges Urteilen abgebildet. Beurteilung versus Wahrnehmung sind schließlich die Präferenzen der letzten Skala.

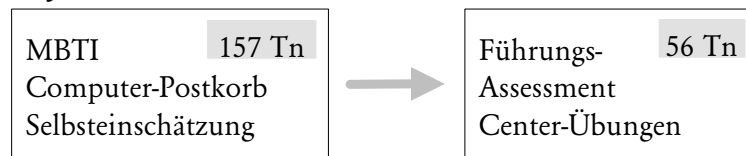
Der Fragebogen besteht aus drei Teilen, die in verschiedener Form insgesamt 90 Forced-choice Items enthalten, wobei es sich um jeweils zwei oder drei alternative Begriffe, Verhaltensweisen, Selbstbeschreibungen und Bewertungen handelt. Bei der Auswertung werden Rohwerte und korrigierte Werte für die Skalen Extraversion Introversion, Sensing, Intuition, Thinking, Feeling, Judging und Perceiving gebildet, die Differenzwerte zwischen den jeweiligen Paaren bestimmt sowie die dazugehörigen Präferenzwerte, welche die Stärke der jeweiligen Typenbevorzugung verdeutlichen, und der Typ des Probanden festgestellt, der schließlich auf jeder der vier Dimensionen eine Präferenzneigung hat. Daraus lassen sich insgesamt 16 Kombinationen bilden.

Zur Vergleichbarkeit des englischsprachigen Originals mit der deutschen Fassung, die hier eingesetzt wurde, werden im Manual keine Angaben gemacht. Daneben gibt es weitere Kritikpunkte an dem Verfahren (Hank & Schwenkmezger, 1992), welche die teststatistische Fundierung des Verfahrens betreffen, sowie die Dokumentation die Konstruktvalidierung und die

Absicherung gegenüber Antworttendenzen. Dennoch wird dieser Persönlichkeitsfragebogen gerade für betriebliche Personalauswahl und -entwicklung häufig eingesetzt.

Der *Fragebogen zur Selbsteinschätzung* wurde von einer Personalberatung entwickelt und zur Verfügung gestellt, der teils offene, teils geschlossene Fragen zu berufliche Vorstellungen und Interessen enthält. Er bezieht sich auf die eigenen Motive, Fähigkeiten und Erfahrungen der Teilnehmer. Die Antworten sollen Rückschlüsse zulassen auf die Ausprägung sozialer, persönlicher und fachlicher Kompetenzen, auf das Führungspotenzial sowie auf persönliches Engagement und Führungsmotivation. Entsprechend beinhaltet der Fragebogen vier Bereiche. Der erste Bereich umfasst berufliche Ziele, Einstellungen und Werte. Im zweiten Abschnitt werden derzeitige Aufgabenschwerpunkte abgefragt. Der dritte Abschnitt behandelt fachliche Kompetenzen, der vierte Führungskompetenz und Führungsmotivation.

1. Jahr



2. Jahr

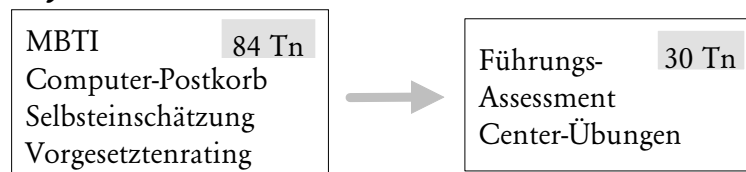


Abbildung 6.1: Ablauf und Stufen der internen Führungskräfteauswahl.

Jeweils der direkte Vorgesetzte eines Teilnehmer bearbeitete nur im zweiten Jahr einen ebenfalls intern entwickeltes *Vorgesetztenrating* für die erste Auswahlstufe. Er enthielt fünf offene Fragen, anhand derer der Vorgesetzte das Verhalten seines Mitarbeiters in führungsrelevanten Situationen sowie dessen Stärken und Schwächen frei beschreiben sollte. Weitere gebundene Items ermöglichten auf einer Ratingskala eine Einschätzung der Kommunikationsfähigkeit, der Teamfähigkeit, der Konfliktfähigkeit und des Führungsverhaltens.

Das *Assessment-Center* für die Teilnehmer der zweiten Auswahlstufe enthielt mehrere Bausteine, Gruppenübungen und Einzelübungen, die als ty-

pisch für Führungssituationen angesehen wurden. Es handelte sich um eine Fallstudie, ein Rollenspiel, ein Konfliktgespräch mit einem fiktiven Mitarbeiter und eine Präsentationsübung, die sich inhaltlich aufeinander bezogen. Die Fallstudie und das Rollenspiel wurden als Gruppenübung durchgeführt, das Konfliktgespräch und Präsentationsübung als Einzelübung. Die Beobachter setzten sich aus Vorgesetzten, Mitarbeitern der Personalabteilung sowie Beratungsmitarbeitern zusammen. Sie bewerteten das Abschneiden der Teilnehmer auf verhaltensverankerten Ratingskalen, die entsprechend der betrieblichen Anforderungsdimensionen zusammengefasst wurden.

Insgesamt nahmen 241 Kandidaten in den beiden aufeinanderfolgenden Jahren an den Verfahren der ersten Stufe teil. Alle Teilnehmer waren Männer. Im ersten Jahr kamen von 157 Teilnehmern an der ersten Auswahlstufe 56 in die zweite Auswahlstufe. Im zweiten Jahr erreichten von 84 Teilnehmern 30 das Assessment-Center. Aus den Ergebnissen der ersten Auswahlstufe wurden nach konjunkten Cutoff-Kriterien von den Beratungsunternehmen und der Personalabteilung die Teilnehmer mit dem Personalvorstand vorgestellt und diskutiert, die in die zweite Auswahlstufe kommen sollten. Diese Vorschläge wurden in der Diskussion größtenteils akzeptiert, im Einzelfall auch korrigiert unter Bezug auf weitere Informationen über die beruflichen Leistungen.

6.3.2 Daten

Die Ergebnisse der Auswahlverfahren wurden in stark aggregierter Form zu vier Zeitpunkten an die wissenschaftliche Begleitung übergeben. Nach Abschluss der ersten Auswahlstufe im ersten Jahr lagen für den Postkorb und die Selbsteinschätzung jeweils die mittleren Ratingwerte vor, die hinsichtlich der fünf Anforderungsdimensionen zusammengefasst waren. Aus dem MBTI lag ein einzelner aggregierter Rating-Wert vor, der die Ausprägung der Dimension Persönlichkeit angeben sollte. Alle Ratingwerte waren von 1 (geringste Ausprägung) bis 7 (höchste Ausprägung) skaliert. Diese elf Werte waren vollständig für alle 157 Teilnehmer vorhanden, außerdem die Angaben über Unternehmenszugehörigkeit und zuständige Urteiler für die Ratings sowie die Information darüber, welche Teilnehmer nach Beratungsvorschlag und nach Unternehmensentscheidung in der zweiten Auswahlstufe teilnehmen sollten. In der zweiten Auswahlstufe lagen für 56 dieser Teilnehmer ebenfalls aggregierte Ratingwerte für die Ausprägung der fünf Anforderungsdimensionen im Assessment-Center vor.

Im zweiten Jahr lagen für die 84 Teilnehmer der ersten Auswahlstufe zunächst wieder jeweils die entsprechenden fünf Ratingwerte für Postkorb und Selbsteinschätzung vor, außerdem fünf Ratingwerte der Vorgesetztenbeur-

teilung, entlang der Anforderungsdimensionen aggregiert. Aus dem MBTI wurden im zweiten Jahr die acht Skalenwerte direkt übergeben. Für diese Teilnehmer war der Beratungsvorschlag und die Unternehmensentscheidung bekannt, wer in der zweiten Auswahlstufe teilnehmen sollten. Von den 30 Teilnehmern der zweiten Stufen gab es aggregierte Ratingwerte für die Ausprägung der fünf Anforderungsdimensionen im Assessment-Center. Für die Rating-Werte wird eine kontinuierliche Abstufung entsprechend der damit abgebildeten Leistung angenommen, weshalb sie bei der statistischen Auswertung als Intervalldaten behandelt wurden.

6.3.3 Evaluationsstrategie

Zunächst war mit den Auftraggebern zu klären, welche Anforderungen sie an die diagnostische Maßnahme stellten und welche weiteren Indikatoren für eine sachgerechte Umsetzung des Verfahrens vorhanden waren. Die Evaluation sollte bis zum Abschluss des Verfahrens und vor dem Hintergrund der verfügbaren Daten erfolgen. So war eine prädiktive Kriteriumsvalidierung des Verfahrens nicht möglich, das naheliegende Analog für eine diagnostische Maßnahme zu einer Impact-Evaluation, es ließ sich nur innerhalb der Maßnahme eine konkurrentes Kriteriumsvalidierung modellieren. Deshalb mussten vorher in Rücksprache mit den Ansprechpartnern der zuständigen Personalabteilung Fragestellungen und Zielkriterien erarbeitet werden.

Neben einer Kriteriumsvalidierung waren deshalb andere Evaluationsstrategien für die Maßnahme erforderlich. Auch im Interesse einer kontinuierlichen Verbesserung des Auswahlverfahrens lag der Schwerpunkt deshalb auf einer normativen Treatment-Evaluation, also Strategien, um festzustellen, ob die Maßnahme tatsächlich so umgesetzt wurde, wie sie intendiert war. Darüber musste die Personalabteilung Auskunft geben. Eine wesentliche Kernannahme bestand darin, dass alle Verfahren alle Anforderungsdimensionen zuverlässig erfassen und abdecken sollten, wobei die Anforderungsdimensionen selbst weitgehend unabhängig voneinander und unabhängig von der Methode sein sollten.

Zudem beruhten etliche Ratingwerte auf Urteiler einschätzung. Die Resultate des Selbsteinschätzungsfragebogens wurden von unterschiedlichen Urteiler n ermittelt. Natürlich sollten diese Werte unabhängig von den Urteiler n sein, es sollte also eine hohe Auswertungsobjektivität gegeben sein. Außerdem sollte der Gesamtwert, also die Empfehlung nach der ersten Stufe unabhängig von den Unternehmen sein, die damit beauftragt waren. Das waren in den beiden Jahren unterschiedliche Unternehmen, die jeweils noch ein Verfahren durchführten. Außerdem wünschte die Personalabteilung Unterstützung bei der Entwicklung eines Entscheidungsmodells.

Als wichtiges Ziel, das eher einer kausativen Impact-Evaluation zuzuordnen ist, waren die am besten Geeigneten zu identifizieren. In Ermangelung eines externen Berufserfolgskriteriums ließ sich die Leistung im Assessment-Center mit den Verfahren der ersten Auswahlstufe korrelieren. Es sollte einen mittleren, aber keinen hohen Zusammenhang geben, da sonst das Assessment-Center redundant wäre. Aus wissenschaftlicher Perspektive ließen sich noch die Vorgesetztenratings als berufliches Erfolgskriterium modellieren. Damit waren mehrer Leitfragen formuliert, die mit unterschiedlichen statistischen Verfahren geprüft werden konnten.

1. **Objektivität der Urteiler** Sind die Ratingwerte aus dem Selbsteinschätzungsfragebogen unabhängig von den zuständigen Urteilern? Da ein Code der Urteiler ebenfalls erfasst worden war, ließ sich ein systematischer Bias zu Lasten einzelner Urteiler über Schätzungen standardisierter Effektgrößen prüfen.
2. **Konstruktvalidität** Sind die Anforderungsdimensionen unabhängig voneinander und unabhängig von der Methode? Dies lässt sich in einer ersten Phase mit einer faktorenanalytischen Multitrait-Multimethod-Validierung prüfen. Nach Rückmeldung der Ergebnisse kann das ideale Modell vor diesem Hintergrund spezifiziert werden und einem plausiblen Modell in konfirmatorischen Faktorenanalysen gegenübergestellt werden. Außerdem kann geprüft werden, ob sich anhand der MBTI-Skalen-Werte die angenommene Faktorenstruktur replizieren lässt.
3. **Kriteriumsvalidität** Lassen sich die Ergebnisse des Assessment-Centers durch das Abschneiden in den Verfahren der ersten Stufe vorhersagen, erfassen also die verschiedenen Methoden die gleichen Konstrukte? In der zweiten Auswahlstufe liegen auch Vorgesetztenratings vor. Diese können als Berufserfolgskriterium modelliert werden. Beide Fragen werden mit Strukturgleichungsanalysen überprüft.
4. **Gewichtung der Instrumente** Erwartet wurde, dass die Ratingwerte aus den Verfahren der ersten Stufe für die Empfehlung zur zweiten Stufe gleichermaßen gewichtet wurden. Dies ließ sich durch einfache Korrelationen der Einzelratings mit der Gesamtempfehlung erfassen. Dies wurde aufgrund unterschiedlicher Verfahren und Zuständiger in den beiden Jahren für beide Jahre separat berechnet.
5. **Entscheidungsmodell** Nachdem sich herausstellte, dass die Ratingwerte in einigen Auswahlstufen und -Verfahren klinisch aggregiert wurden, sollte alternativ ein statistisches Entscheidungsmodell erarbeitet und umgesetzt werden. Die Modellierung wird kurz skizziert.

Wie im folgenden Ergebnisteil erläutert, wurden diese Analysen im Verlauf von den zwei Jahren mehrfach gerechnet und zurückgemeldet, verbunden mit einer Diskussion, wie die Diagnostik optimiert werden konnte. Nach der Grobeinteilung von Scriven (1991) handelt es sich damit auch um eine formative Evaluation, was zeigt, dass dies nicht allein qualitative Methoden impliziert.

6.4 Ergebnisse

Nachdem die Fragestellungen in einer qualitativen Phase gemeinsam mit den Auftraggebern erarbeitet wurden, konnte die quantitative Auswertung stattfinden. Diese Ergebnisse meldete die wissenschaftliche Begleitung die Ergebnisse zu vier Zeitpunkten zurück, immer nach Abschluss einer Auswahlstufe. Die Ergebnisdarstellung erfolgt hier zusammengefasst nach den Fragestellungen, die Konsequenzen der Feedbacks zu den vier Zeitpunkten für das weitere Verfahren werden in einem eigenen Teilkapitel berichtet.

6.4.1 Objektivität der Urteiler

In der ersten Auswahlstufe war der Code der sieben Urteiler bekannt, welche die Ratingwerte des Selbsteinschätzungsfragebogens bestimmt hatten. Es handelte sich um fünf Ratingwerte, welche die Ausprägungen jedes Teilnehmers bei der Bearbeitung des Fragebogens auf den fünf Anforderungsdimensionen angaben. Bekannte Urteilertendenzen sind die persönliche Neigung zu besonders günstigen, oder besonders ungünstigen Werten, der sogenannte Milde- und umgekehrt der Strenge-Effekt. Es ließe sich varianzanalytisch prüfen, ob ein systematischer Effekt durch die Urteiler auf die Ausprägungen vorhanden war, also eine Verletzung der Auswertungsobjektivität bestand. Die Auswertungsobjektivität lässt sich als ein erster Indikator für eine Treatment-Evaluation nutzen.

Dazu könnte man eine Varianzanalysen durchführen, mit den Ratingwerten als abhängigen Variablen, der Faktor wäre die Zuordnung der Urteiler mit fünf Stufen in Form der Urteiler, wobei es sich hier um einen randomisierten Faktor handelt. Allerdings waren die Voraussetzungen für eine multivariate oder eine univariate Varianzanalyse verletzt. Zwar kann man den Rating-Daten Intervallskalenniveau unterstellen, doch ist weder eine Normalverteilung der abhängigen Variablen gegeben, noch Varianzhomogenität, was auf das dritte Problem, stark unterschiedliche Zellengrößen zurückzuführen ist. Die sieben Urteiler hatten unterschiedlich viele Teilnehmer bewertet (siehe Tabelle 6.1), was sich anders als bei einer Laborerhebung in der Feldsituati-

on nicht im Sinne dieser Auswertungsfrage optimieren lässt. Auch die Power leidet darunter. Damit ist die Interpretation einer varianzanalytischen Auswertung sehr fraglich.

Urteiler Code (N)	Persön- lichkeit	System. Denken	Soziale Kompet.	Selbst- managem.	Untern. Handeln	Urteiler- tendenz
1 (19)	-.01	-.06	.04	-.07	.12	.00
2 (71)	.15	.07	.06	.10	.15	.11
3 (21)	.01	.07	-.01	.07	.00	.03
4 (9)	-.05	-.04	-.18	-.05	-.11	-.09
5 (8)	.15	-.04	.11	.02	.03	.05
6 (18)	-.11	.00	-.01	-.05	-.09	-.05
7 (9)	-.15	.00	-.01	-.03	-.11	-.06

Tabelle 6.1: Effektstärken für die Urteilertendenzen beim Selbsteinschätzungsfragebogen auf fünf Dimensionen (N=155).

Um dennoch eine Aussage über potenzielle Urteilertendenzen treffen zu können, wurden Effektstärken nach Cohen (1988) berechnet, also die standardisierte Abweichung der mittleren Ratings eines Urteilers vom Gesamtmittelwert für jede Ratingvariable. Außerdem ließ für jeden Urteiler über die fünf Variablen hinweg die Effektstärke seiner mittleren Abweichung berechnen, also ein aggregiertes Maß für seine Urteilertendenz. Cohen unterscheidet in seiner Einteilung konventionalisierter Kriterien schwache, mittlere und starke Effekte. Ein schwacher Effekt liegt unterhalb $d = 0.20$.

Die Ergebnisse in Tabelle 6.1 zeigen nach Cohen (1988) durchgängig nur schwache Effekte, die noch deutlich unter der Höchstgrenze für schwache Effekte nach der Einteilung von Cohen liegen. Selbst wenn man die Abweichungsmaße für die einzelnen Urteiler über die verschiedenen Dimensionen aggregiert, um ein stabileres Maß für die individuellen Tendenzen zu bekommen, sind nur schwache Effekte zu beobachten. Das bedeutet, dass die Urteiler ihre Einschätzungen weder durch Strenge- noch durch Milde-Effekte verzerrten. An die Auftraggeber lässt sich also rückmelden, dass die verschiedenen Urteiler die verschiedenen Teilnehmer insofern fair behandelten, dass keine individuellen Tendenzen in die Ratingwerte des Selbsteinschätzungsfragebogens mit einfließen.

6.4.2 Konstruktvalidität

Eine wesentliche Frage für die Auftraggeber war, inwieweit die Managementdiagnostik entsprechend den inhaltlichen Erwartungen umgesetzt wurde, es

sollten also Strategien zur Treatment-Evaluation generiert werden. Die Erwartungen bestanden darin, dass die betrieblich definierten Anforderungsdimensionen einerseits unabhängig voneinander seien, da sie ja auch inhaltlich unterschiedlich definiert waren. Weiterhin sollten alle Verfahren diese Anforderungsdimensionen vergleichbar erfassen. Damit sind Aspekte der Konstruktvalidität angesprochen, und die Datenlage bildet die typische Voraussetzung für eine Multitrait-Multimethod-Validierung, da die Messungen mehrerer Konstrukte durch verschiedene Verfahren vorliegen, und mit jeder Methode mehrere Konstrukte erfasst wurden, ähnlich wie beim Assessment-Center.

Faktor (Varianzanteil)	1 (40%)	2 (38%)
Selbsteinschätzung		
Persönlichkeit	.85	.31
System. Denken	.81	.34
Soziale Kompetenz	.82	.27
Selbstmanagement	.87	.24
Untern. Handeln	.86	.32
Computer-Postkorb		
Persönlichkeit	.26	.88
System. Denken	.23	.89
Soziale Kompetenz	.41	.64
Selbstmanagement	.32	.77
Untern. Handeln	.29	.88

Tabelle 6.2: Faktorladungen der Ratingwerte im ersten Jahr nach Varimax-Rotation, Extraktion nach Kaiser-Kriterium (N=157).

Eine Multitrait-Multimethod-Validierung lässt sich faktorenanalytisch oder korrelativ durchführen. Da die Stichprobe groß genug ist, wurden für die beiden Jahre Faktorenanalysen durchgeführt. Im ersten Jahr lagen in der ersten Stufe zwei Verfahren vor, die alle Anforderungsdimensionen abdecken, sowie das Assessment-Center in der zweiten Stufe. Im zweiten Jahr wurde in der ersten Auswahlstufe noch ein drittes Instrument hinzugenommen, das Vorgesetztenurteil, ebenfalls mit Ratingwerten auf allen Dimensionen. Dabei wurden im ersten Jahr alle 157 Teilnehmer einbezogen und 84 Teilnehmern im zweiten Jahr, wobei von 66 Teilnehmern die kompletten Daten vorlagen, so dass die restlichen aufgrund von fehlenden Daten zeilenweise beziehungsweise fallweise ausgeschlossen wurden.

Zum einen wurde explorativ eine Hauptkomponentenanalyse gerechnet, um zu sehen, wie viele Faktoren überhaupt extrahiert wurden, und diese nach einer Varimax-Rotation interpretiert. Zum anderen wurden für diese Varimax-Rotation fünf Faktoren vorgegeben, auf denen jeweils die Rating-

Faktor (Varianzanteil)	1 (28%)	2 (25%)	3 (24%)
Selbsteinschätzung			
Persönlichkeit	.91	.11	.15
System. Denken	.85	.07	.27
Soziale Kompetenz	.90	.11	.28
Selbstmanagement	.95	.07	.04
Untern. Handeln	.90	.12	.24
Vorgesetztenurteil			
Persönlichkeit	.12	.89	.03
System. Denken	.14	.86	.17
Soziale Kompetenz	.05	.76	.18
Selbstmanagement	.07	.85	.03
Untern. Handeln	.04	.88	.16
Computer-Postkorb			
Persönlichkeit	.26	.17	.80
System. Denken	.08	.09	.87
Soziale Kompetenz	.19	.02	.71
Selbstmanagement	.15	.10	.78
Untern. Handeln	.21	.23	.86

Tabelle 6.3: Faktorladungen der Ratingwerte im zweiten Jahr nach Varimax-Rotation, Extraktion nach Kaiser-Kriterium (N=66).

werte für die Anforderungsdimensionen hoch laden sollten, um festzustellen, ob sich dieses Muster tatsächlich zeigt. Tabelle 6.2 und Tabelle 6.3 zeigen die Ergebnisse der explorativen Faktorenanalyse für die Varimax-Rotation. Nach dem Kaiser-Kriterium ließen sich jeweils so viele Faktoren extrahieren, wie Verfahren vorhanden waren, also zwei Faktoren im ersten Jahr und drei Faktoren im zweiten Jahr, die auch eine hohe Varianzaufklärung erzielten. Die Faktoren waren anhand ihrer Ladungen klar als Methodenfaktoren zu interpretieren.

Im ersten Jahr klärte der erste Faktor der explorativen Faktorenanalyse hauptsächlich die Varianz auf, die auf die Ratingwerte des Selbsteinschätzungsfragebogens zurückgeht, der zweite Faktor repräsentiert den computer-gestützten Postkorb. Im zweiten Jahr war das Muster ähnlich, wobei sich ein weiterer Faktor analog als Faktor für das Vorgesetztenurteil interpretieren ließ. Damit war weder eine diskriminante Validität gegeben, also geringe Korrelationen der Ratingwerte für unterschiedliche Konstrukte bei gleichen Methoden, noch eine konvergente Validität, also hohe Korrelationen der Ratingwerte gleicher Konstrukte der unterschiedlichen Methoden. Andernfalls würde die Faktorenanalyse in beiden Fällen fünf Faktoren extrahieren, auf denen die zwei oder drei Ratingwerte der gleichen Dimensionen jeweils auf einem Faktor hoch und auf den anderen gering laden.

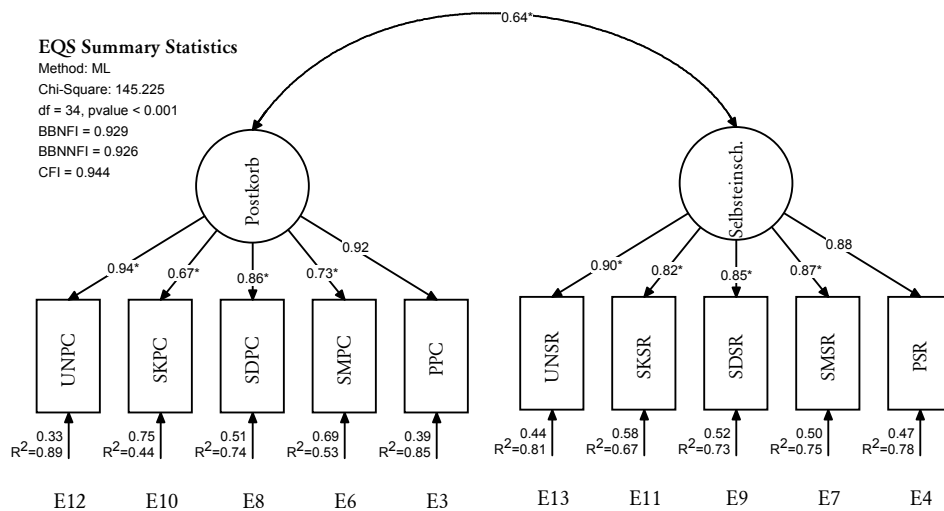


Abbildung 6.2: Konfirmatorische Faktorenanalyse für ein angenommenes Urteilsmodell anhand der Methoden. N = 157, Teilnehmer im ersten Jahr. Berechnet mit EQS (Bentler, 1998).

Wenn also diese Struktur in den Daten repräsentiert ist, wie die Auftraggeber annahmen, sollte sich diese explorativ oder hypothesentestend aufdecken lassen, möglicherweise mit anderen Auswertungsstrategien. Bei der Varimax-Rotation mit fünf vorgegebenen Faktoren ließen sich die ersten drei Faktoren wieder als Methodeneffekte interpretieren, und kein Faktor im Sinne der inhaltlichen Konstrukte. Außerdem wurde eine schiefwinklige Faktorenanalyse, zunächst explorativ und anschließend nach Vorgabe von fünf zu extrahierenden Faktoren gerechnet. Explorativ resultierten daraus wieder die Methodenfaktoren, nur interkorreliert. Auch mit Vorgabe von fünf Faktoren konnte kein einziger Faktor im Sinne der Anforderungsdimensionen interpretiert werden.

Um dennoch eine dimensionale Struktur aufzudecken, die durch die Methodenfaktoren maskiert wird, wurde schließlich eine Faktorenanalyse der Residuen nach Extraktion der Methodenfaktoren durchgeführt, sowohl für die Residuen nach rechtwinkliger als auch für die Residuen aus der schiefwinkligen Rotation. Auch dies führte zu keinem Ergebnis im Sinne der formulierten Hypothese. Aus Platzgründen sind die Ergebnisse hier nicht abgebildet.

Da die strukturentdeckenden Herangehensweisen in keinem Fall die Dimensionalitätshypothese stützten, wurde zur abschließenden Klärung eine hypothesentestende Vorgehensweise gewählt, in der zwei alternative Modelle in einer konfirmatorischen Faktorenanalyse daraufhin verglichen wurden, wie

gut sie die empirische Datenlage abdecken. Nach Rücksprache mit den Auftraggebern standen zwei Varianten zur Debatte: Ein Modell mit fünf unabhängigen Faktoren im Sinne der Anforderungsdimensionen sowie ein Modell mit interkorrelierten Methodenfaktoren.

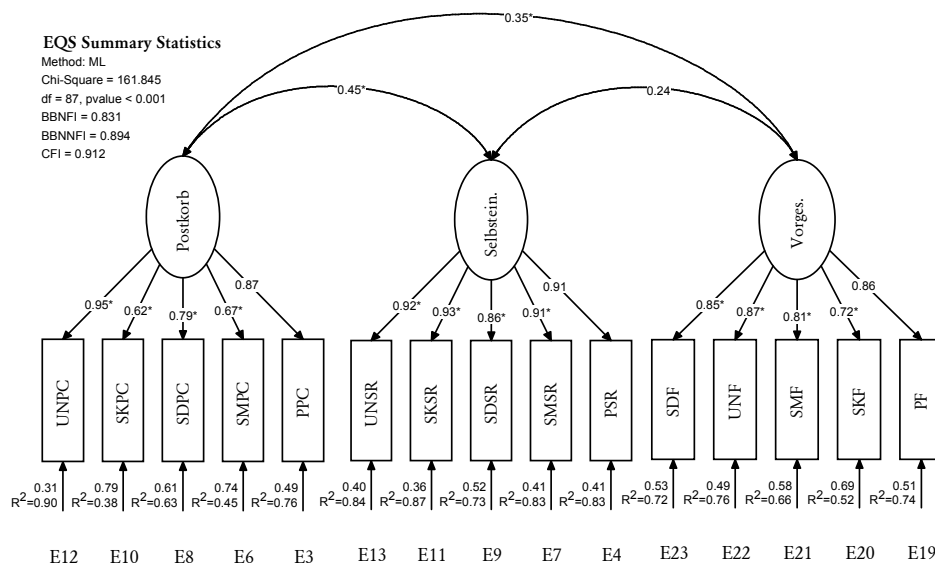


Abbildung 6.3: Konfirmatorische Faktorenanalyse für ein angenommenes Urteilsmodell anhand der Methoden. N = 66, Teilnehmer im zweiten Jahr. Berechnet mit EQS (Bentler, 1998).

Was sich bereits bei den explorativen Faktorenanalysen abgezeichnet hatte, wurde hier deutlich. Das nicht intendierte Urteilsmodell, bei dem die Ratingwerte vor allem auf Methodenfaktoren hoch laden, konnte die Daten wesentlich besser abbilden als das eigentlich angenommene. Abbildung 6.2 und Abbildung 6.3 zeigen die erfolgreichereren, aber nicht intendierten Modelle, welche die Ratings als manifeste Variablen den Methoden als latente Variablen oder Faktoren zuordnen, sowie die standardisierten Parameter nach Berechnung in EQS (Bentler, 1998). Die idealen Modelle mit fünf unabhängigen Dimensionen, auf denen die zugehörigen Ratingwerte hoch laden sollten, wurden ebenfalls gerechnet, sind jedoch aus Platzgründen nicht abgebildet.

Im ersten Jahr betrug der Model-Fit für das gewünschte Urteilsmodell $CFI = 0.299$, dagegen erzielte das nicht intendierte Modell mit den zwei Methodenfaktoren, das die explorative Faktorenanalyse nahe legt, eine sehr gute Modellanpassung mit $CFI = 0.944$. Der Model-Fit für das ideale Urteilsmodell war im zweiten Jahr sogar noch geringer mit $CFI = 0.077$, das explorativ

vermutete Urteilsmodell stellte jedoch wieder eine gute Modellanpassung dar mit $CFI = 0.912$. Im zweiten Jahr war die Stabilität der Strukturgleichungsmodelle angesichts der Stichprobe von 66 Teilnehmern fraglich, es ist jedoch zu erwarten, dass sich angesichts des guten Modell-Fits bei einer größeren Stichprobe ähnliche Muster ergeben.

Die beschriebenen Analysen hatten also das Ziel, die Urteilsbildung der eingesetzten Rater zu modellieren und zu prüfen, ob sie das von den Auftraggebern vorgestellte nomologische Netzwerk abbildet. Die Ergebnisse, die sogar in zwei aufeinanderfolgenden Jahren in die gleiche Richtung wiesen, legen nahe, dass die Urteiler einem Teilnehmer innerhalb eines Verfahrens eher ähnliche Ratings gaben, als innerhalb einer Dimension. Dieses Phänomen könnte man als Methodenartefakt bezeichnen, dabei ähnelt es auch dem Halo-Effekt, da Ausprägung einer Dimension mit den Ausprägungen anderer Dimensionen eines Verfahrens korreliert sind.

Als Erklärung könnten Defizite bei der Urteilerschulung vermutet werden, ebenso eine unpräzise Operationalisierung und verbale Verankerung der Ratingwerte für die Dimensionen bei den jeweiligen Verfahren. Weiterhin ist es möglicherweise eine unrealistische Annahme, die inhaltlich sehr unterschiedlichen Diagnoseverfahren könnten in vergleichbarer Weise die Anforderungsdimensionen erfassen. Zum Beispiel mussten die Indikatoren für soziale Kompetenz beim Postkorb andere sein als bei der Selbsteinschätzung oder beim Vorgesetztenrating. Vielleicht bilden schließlich die Anforderungsdimensionen selbst, deren Inhalt und Beziehung sich zunächst als Hypothese auffassen lassen, die Realität nicht in dieser Form ab. Deshalb lässt sich nur konstatieren, dass eine Diskrepanz zwischen gewünschtem und realisiertem Urteilsmodell besteht.

Schließlich lagen noch die acht Skalenwerte des MBTI vor, leider nicht die einzelnen Itemwerte. Da der MBTI in der Literatur umstritten ist, sollte eine Faktorenanalyse prüfen, ob sich die Skalen und ihre angenommenen Zusammenhänge replizieren ließen. Im Testmanual (Briggs & Briggs-Myers, 1991) wird zunächst berichtet, dass die acht Subskalen und auch die vier Hauptskalen interkorreliert seien, sich jedoch mit einer Varimax-Rotation ebenfalls die vier Subskalen replizieren ließen. Demnach sollten immer die beiden zusammengehörigen Skalen wie Extraversion und Introversion hoch positiv und negativ auf einem Faktor laden, und außerdem für jede Skala ein Faktor extrahierbar sein.

Tabelle 6.4 zeigt die Faktorladungen der Subskalen nach Varimax-Extraktion von vier Faktoren. Nach dem Kaiser-Kriterium wurden drei Faktoren extrahiert, der Scree-Plot legt jedoch auch die Extraktion von vier Faktoren nahe (vergleiche Abbildung 6.4). Anhand der Faktorladungen ließen sich zumindest die Subskalen im Sinne der vier Skalen interpretieren.

Faktor (Varianzanteil)	1 (24%)	2 (24%)	3 (23%)	4 (23%)
Sensing	-.94	-.08	.06	-.26
Intuition	.95	.07	.02	.23
Extraversion	.04	.98	-.01	.01
Introversion	-.09	-.98	-.01	-.03
Thinking	-.01	.03	.93	-.22
Feeling	.03	.04	-.95	.15
Judging	-.30	-.09	.20	-.91
Perceiving	.24	.05	-.23	.92

Tabelle 6.4: Faktorstruktur der MBTI-Subskalenwerte im zweiten Jahr nach Varimax-Rotation, Extraktion nach Scree-Kriterium (N=79).

Allerdings gab es im Auswahlverfahren keine Strategie, die Skalen oder Subskalen den Anforderungsdimensionen zuzuordnen, sie fallen damit aus dem diagnostischen Rahmenmodell heraus.

Auf Nachfrage gab die zuständige Beratung an, dass die MBTI-Werte bei besonderen Fällen ergänzend für die Ratingwerte des Selbsteinschätzungsfragebogens hinzugezogen wurden, also eine optionale Rolle spielten. Welche Fälle und welche Anforderungsdimensionen dies waren, und in welcher Form die MBTI-Ergebnisse hinzugezogen wurden, war nicht dokumentiert und ließ sich im Nachhinein nicht mehr ermitteln. Zumindest zeigten sich keine signifikanten korrelativen Zusammenhänge mit den Ratingwerten des Selbsteinschätzungsfragebogens, auch die Effektstärken für die Korrelationskoeffizienten waren nach Cohen (1988) gering.

Man kann wohl davon ausgehen, dass die Testdurchführung und Auswertung sachgerecht erfolgte, auch wenn sie keine hohe Relevanz im Gesamtverfahren hatte. Das mit der Datenintegration beauftragte Unternehmen hatte die MBTI-Werte nicht zur Verfügung, damit spielten sie keine Rolle bei der Empfehlung für die zweite Auswahlstufe. Weiterhin gab es keine signifikanten Zusammenhänge mit den Ergebnissen in den anderen Verfahren, auch nicht mit der Empfehlung für das Assessment-Center. Deshalb sind in den folgenden Abschnitten Befunde zum MBTI nicht mehr aufgeführt.

6.4.3 Kriteriumsvalidität

Ursprünglich wurden Validierungsstudien als wesentliche Möglichkeit gesehen, eignungsdiagnostische Maßnahmen zu evaluieren. Damit sind Aspekte einer diagnostischen Impact-Evaluation angesprochen. Die Kriteriumsvalidität soll anhand der Korrelation mit äußeren Indikatoren absichern, dass mit den zu validierenden Verfahren Vergleichbares erfasst wird.

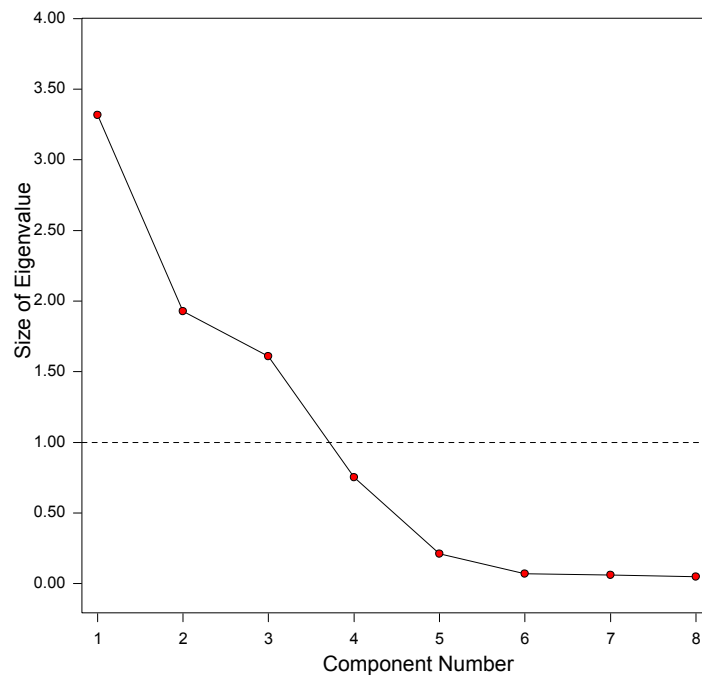


Abbildung 6.4: Scree-Plot für die Faktorenanalyse der Subskalenwerte des MBTI. Berechnet mit EQS (Bentler, 1998).

In der vorliegenden Studie lassen sich zwei Hinweise auf Kriteriumsvalidität berechnen. Zum einen kann die Erwartung des Auftraggebers abgebildet werden, dass die Verfahren der ersten Auswahlstufe Vergleichbares erfassen wie das Assessment-Center der zweiten Auswahlstufe. Zum anderen könnte aus wissenschaftlicher Perspektive das Vorgesetztenurteil als Berufserfolgskriterium für die anderen Verfahren als Prädiktoren modelliert werden. Das entspricht nicht der Logik des Urteilsprozesses aus der Sicht des Auftraggebers, wo das Vorgesetztenurteil einen Prädiktor darstellt, aber es ist in den meisten Validierungsstudien üblich und liefert zusätzliche Informationen über die Prognosegüte der restlichen Verfahren hinsichtlich eines verbreiteten Berufserfolgskriteriums.

Zunächst wurde versucht, mit Hilfe eines Strukturgleichungsmodells die Leistung im Assessment-Center als Kriterium oder endogene Variable durch die erzielten Werte in den beiden Verfahren Postkorb und Selbsteinschätzungsfragebogen als endogene Variablen vorherzusagen. Alle drei Verfahren waren als latente Variablen abgebildet. Die zugehörigen manifesten Variablen bestanden in den Ratingwerten für die Teilleistungen im jeweiligen Verfah-

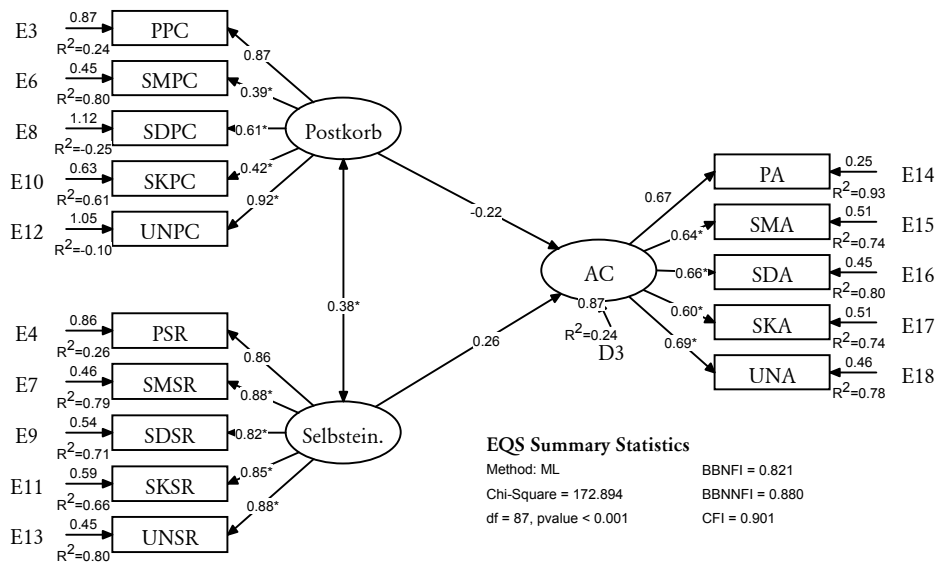


Abbildung 6.5: Strukturgleichungsmodell zur Kriteriumsvalidierung zweier Verfahren am Assessment-Center als endogene Variable. $N = 78$. Berechnet mit EQS (Bentler, 1998).

ren. Das Vorgesetztenurteil wurde ausgeklammert, da sonst die Stichprobe zu klein würde. Analog zu den Ergebnissen aus dem vorherigen Abschnitt wurde eine Abhängigkeit zwischen den beiden Prädiktoren angenommen.

Wie in Abbildung 6.5 zu sehen ist, erzielte dies insgesamt eine gute Modellanpassung mit einem Model-Fit von $CFI = 0.901$, jedoch waren die beiden Pfadkoeffizienten der direkten Pfade vom Postkorb und vom Selbsteinschätzungsfragebogen nicht signifikant. Außerdem korrelierten der Postkorb als latente Variable negativ mit den Assessment-Center-Ergebnissen. Sicherlich ist dies mit durch eine Einschränkung der Prädiktorvarianz zu erklären, da nur die ausgewählten Teilnehmer, mit 78 Personen ungefähr ein Drittel der Gesamtstichprobe, am Assessment-Center teilnahmen, welche die besten Resultate in der ersten Auswahlstufe erzielten.

Von einer eingeschränkten Prädiktorvarianz kann bei der zweiten Auswertung (Abbildung 6.6) nicht ausgegangen werden. Hier wurde das Vorgesetztenurteil als Berufserfolgskriterium und endogene latente Variable modelliert, und wiederum der Postkorb und der Selbsteinschätzungsfragebogen als exogene Variablen. Auch hier war die gesamte Modellpassung gut mit einem Model-Fit von $CFI = 0.912$, wobei die Interkorrelation zwischen den beiden exogenen latenten Variablen höher war als die beiden Pfadkoeffizien-

ten auf die endogene latente Variable. Immerhin war die Partialkorrelation der Postkorb-Übung mit dem Vorgesetztenurteil signifikant und positiv und betrug $r = 0.31$. Außerhalb des Strukturgleichungsmodells wurde noch die Partialkorrelation der Faktorwerte des Assessment-Centers mit dem Vorgesetztenurteil berechnet. Sie betrug $r = 0.14$ für $n = 27$.

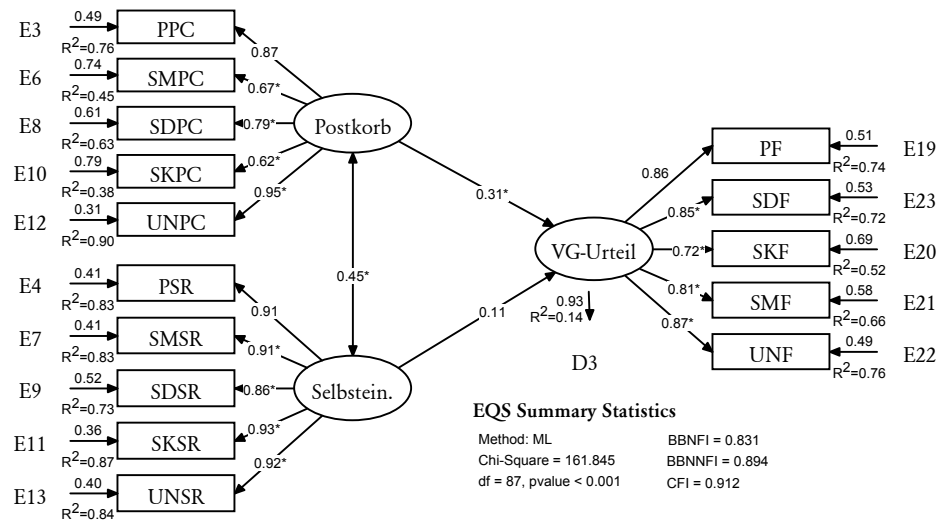


Abbildung 6.6: Strukturgleichungsmodell zur Kriteriumsvalidierung zweier Verfahren am Vorgesetztenurteil als endogene Variable. $N = 66$. Berechnet mit EQS (Bentler, 1998).

Geht man davon aus, dass die computergestützte Postkorb-Übung Anforderungen an komplexes Problemlösen stellt, beispielsweise Prioritätensetzung und Maßnahmenplanung, während der Selbsteinschätzungsfragebogen auf eher motivationale Aspekte abzielt, lässt sich dies als Hinweis auf die berufliche Relevanz von Problemlösefähigkeit interpretieren. Insgesamt erscheint die Kriteriumsvalidität der diagnostischen Maßnahme optimierbar. Dabei sind neben der eingeschränkten Prädiktorvarianz im ersten Modell mehrere Erklärungen beziehungsweise Ansatzpunkte denkbar. Es könnte am Urteilsprozess selbst liegen, die Operationalisierung war nicht angemessen, oder die Ratingwerte sind zu stark aggregiert.

6.4.4 Gewichtung der Instrumente

Etwa ein Drittel der Kandidaten gelangte aufgrund des Abschneidens von der ersten Auswahlstufe in die zweiten Auswahlstufe, die aus dem Assessment-

Center bestand. Um diese Auswahl zu treffen, integrierte eines der beteiligten Beratungsunternehmen die angefallenen Daten und gab eine dichotome Empfehlung ab, das heißt, es empfahl entweder einen Kandidaten für die nächste Stufe oder lehnte ihn ab. Im ersten Jahr gab zunächst eine Beratung eine Empfehlung ab, die noch in der Konzernzentrale diskutiert und teilweise modifiziert wurde, so dass die Information über die ausgesprochene Empfehlung sowie die letztendliche Teilnahme an der zweiten Stufe zur Verfügung gestellt wurde. Im zweiten Jahr hatte eine andere Beratung diese Integration in der Hand und war direkt in die Diskussion mit der Konzernzentrale eingebunden. Davon stand nur die Information über die Teilnahme am Assessment-Center zur Verfügung.

Für jeden Kandidaten gab es im ersten Jahr insgesamt elf Ratingwerte, jeweils fünf aus dem Selbsteinschätzungsfragebogen, und weitere fünf aus dem computergestützten Postkorb sowie ein Wert aus dem MBTI. Im zweiten Jahr lagen für jeden Kandidaten insgesamt 15 Ratingwerte vor, wie im Vorjahr jeweils fünf aus dem computergestützten Postkorb und dem Selbsteinschätzungsfragebogen. Zusätzlich wurde im zweiten Jahr ein Vorgesetztenrating hinzugenommen.

	Empfehlung	Teilnahme
Computer-Postkorb	.45	.29
Selbsteinschätzung	.64	.33
P der Differenz (Fishers Z-Wert)	.02	.37

Tabelle 6.5: Modellierung der Auswahlentscheidungen im ersten Jahr. $N = 157$, für alle Korrelationen $p < 0.001$.

Nun interessierten sich die Auftraggeber für eine Modellierung des Urteilsprozesses bei der endgültigen Gewichtung der Verfahren. Dazu wurden die Ratingwerte der verschiedenen Verfahren mit der ausgesprochenen Empfehlung sowie der tatsächlichen Teilnahme korreliert. Zu diesem Zweck wurden Faktorwerte auf der Basis der oben beschriebenen Varimax-Rotation berechnet und diese mit der Empfehlung (nur im ersten Jahr) und der Teilnahme korreliert. Es wurden die Faktorwerte aus der orthogonalen Rotation genutzt, um die jeweils unabhängigen Varianzanteile mit der Teilnahmeempfehlung und der tatsächlichen Teilnahme zu korrelieren. Daraus resultierten immer punkt-biseriale Korrelationskoeffizienten.

Tabelle 6.5 zeigt für das erste Jahr, dass die faktorenanalytisch aggregierten Ratingwerte der Selbsteinschätzung signifikant höher als der computergestützte Postkorb mit der Assessment-Center-Empfehlung korrelierte.

Die Unternehmensberatung, die den Postkorb durchgeführt hatte, war auch für die Integration der Informationen zur Teilnahmeempfehlung zuständig. Diese Empfehlung bildete die Diskussionsbasis für die Entscheidung des Unternehmens, die Teilnehmer in das Assessment-Center zu schicken. Die Teilnahmeempfehlung und die tatsächliche Teilnahme korrelierten zu $r = 0.40$ miteinander.

Berechnet man nun die Korrelationen der beiden Verfahren mit der tatsächlichen Teilnahme, sinken nicht nur beide Korrelationskoeffizienten, sondern es besteht auch kein signifikanter Unterschied mehr zu Gunsten des Selbsteinschätzungsfragebogens. Das bedeutet, von Seiten des Unternehmens wurden wahrscheinlich weitere Informationen herangezogen, die nicht mit den beiden Verfahren abgedeckt sind, was außerdem die Tendenz der für die Selbsteinschätzung zuständigen Beratung ausgleicht, ihr eigenes Verfahren deutlich höher zu gewichten, als das der „Konkurrenz“.

Im zweiten Jahr übertrug das Unternehmen die Datenintegration der Beratung, die für den computergestützten Postkorb zuständig war. Hier gab es kein anschließendes Korrektiv durch das Unternehmen, sondern das Unternehmen legte seine abschließende Entscheidung in der Diskussion mit der Beratung fest. Nur die Informationen über die tatsächliche Teilnahme an der zweiten Auswahlstufe wurde deshalb zur Verfügung gestellt. Das Vorgesetztenurteil führte die zentrale Personalabteilung unternehmensintern durch und gab die Ergebnisse an die Beratung weiter.

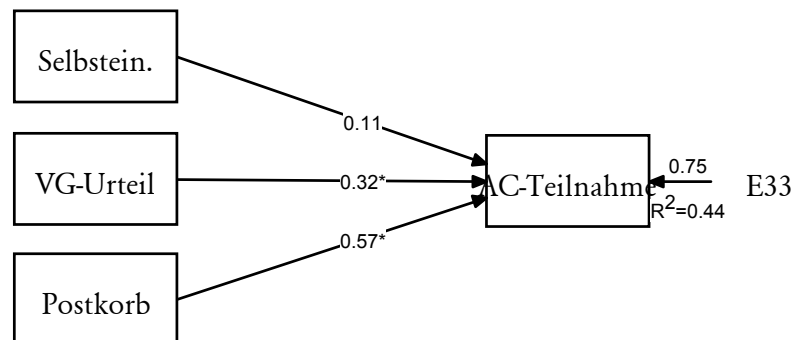


Abbildung 6.7: Modellierung der Auswahlentscheidungen im zweiten Jahr. $N = 66$. Berechnet mit EQS (Bentler, 1998).

Abbildung 6.7 zeigt die punkt-biserialen Korrelationen der unabhängigen Faktorwerte für die drei eingesetzten Verfahren mit der AC-Teilnahme. Am höchsten korrelierte der Postkorb mit $r = 0.57$, $p < 0.001$ mit der AC-

Teilnahme, gefolgt von dem Vorgesetztenurteil mit immer noch $r = 0.32$, $p < 0.05$ und der geringen, nicht signifikanten Korrelation des Selbsteinschätzungsfragebogens mit $r = 0.11$. Auch hier wurde aufgrund Fishers Z-Werten die Signifikanz der Korrelationsdifferenzen berechnet. Dabei war lediglich der Unterschied zwischen den Korrelationen des Postkorbs und der Selbsteinschätzung mit der Teilnahme signifikant ($p < 0.01$), die anderen beiden waren bei dieser Stichprobengröße nicht mehr signifikant für $p < 0.05$.

Es lässt sich also für das zweite Jahr konstatieren, dass die Informationen aus den drei Auswahlverfahren unterschiedlich gewichtet in die Entscheidung für die zweite Auswahlstufe einfließen, am höchsten gewichtete dabei die zuständige Unternehmensberatung ihr eigenes Verfahren, uns am geringsten das der anderen Beratung. Es war zu vermuten, dass das Unternehmen das so nicht intendiert hatte, da es ja die beiden Beratungen gleichberechtigt engagiert und vergütet hatte. Deshalb wurden diese Ergebnisse ohne weiteren Kommentar im Rahmen der Evaluation rückgemeldet, um dem Unternehmen die Gelegenheit zu geben, Rückfragen zu stellen und bei Bedarf Kommentare und Empfehlungen der wissenschaftlichen Begleitung einzuholen.

6.4.5 Entscheidungsmodell

Das Unternehmen betrachtete mehrere Evaluationsergebnisse als erwartungswidrig. Insbesondere die Ergebnisse zur Dimensionalität der Ratingwerte und die unterschiedliche Gewichtung der Verfahren von der ersten zur zweiten Urteilsstufe waren so nicht gewünscht. Deshalb befragten die Unternehmensvertreter die wissenschaftliche Begleitung nach Erklärungen und Optimierungsmöglichkeiten, ohne den Gesamtprozess erheblich zu verändern, was sich auch aus politischen Gründen verbot. Es ging also darum, im Sinne einer Treatment-Evaluation auf formativem Unterstützung bei der Gestaltung des Urteilsprozesses einzuholen, nachdem dieser vorher summativ evaluiert worden war.

Zunächst wurden die Auftraggeber in diesem Kontext über die Bedeutung klinischer und statistischer Urteilsmodelle informiert sowie deren Konsequenzen. Auch war noch offen, wie die Auswahl in der zweiten Auswahlstufe operationalisiert werden sollte. Deshalb kamen die Auftraggeber zum Schluss, dass sie für das Assessment-Center im zweiten Jahr ein statistisches Entscheidungsmodell wünschten, um die Objektivität in diesem Schritt zu optimieren und die Gesamtbewertung unabhängig von den damit Beauftragten zu machen. Es konnte den Auftraggebern deutlich gemacht werden, dass ein statistisches Urteilsmodell durchaus inhaltliche Erwägungen berücksichtigen kann und soll, aber sicherstellt, dass dies für alle Teilnehmer gleichermaßen angewendet wird.

Deshalb musste zunächst in Abhängigkeit von den Vorstellungen der Auftraggeber ein solches Modell spezifiziert werden. Dabei waren mehrere Punkte in mehreren Diskussionsrunden zu entscheiden. Ziel der Auftraggeber war, die Teilnehmer letzten Endes nach ihrem Abschneiden einem von vier Leistungskategorien zuzuordnen. Dabei definierten sie diese Leistungskategorien folgendermaßen: A als höchste Leistungskategorie stand für *High Potential*, Teilnehmer der Stufe B hatten *Geschäftsführerpotenzial*, C stand für *Eingeschränktes Geschäftsführerpotenzial* und bei D war *kein unternehmerisches Potenzial* gegeben. Die Teilnehmer mit den Potenzialstufen A und B sollten gezielt weitergefördert werden, die anderen Teilnehmer erhielten eine Beratung hinsichtlich alternativer Perspektiven, zum Beispiel Fachkarriere.

Sehr schnell waren sich die Auftraggeber einig, dass sie alle fünf Anforderungsdimensionen gleich gewichten wollten. Außerdem musste entschieden werden, wie mit unterschiedlichen Ausprägungen eines Teilnehmers auf den Anforderungsdimensionen umgegangen werden sollte. Zunächst hatten die Auftraggeber die Vorstellung, multiple Cutoffs festzulegen, in der Form, dass eine Leistungskategorie dann vergeben wurde, wenn in allen Dimensionen mindestens eine bestimmte Ausprägung vorhanden war. Außerdem musste entschieden werden, bei welcher Ausprägung welche Potenzialstufe gegeben war. Dies konnte entweder kriteriumsorientiert oder normorientiert an der Stichprobe festgelegt werden.

Schließlich entschieden sich die Auftraggeber dafür, zunächst ein arithmetisches Mittel aller fünf Ratingwerte zu bilden, und für diesen Gesamtwert Prozentränge zu bilden. Dabei sollten Teilnehmer mit dem Prozentrang 100 - 85 der Kategorie A zugewiesen werden, für Prozentrang 84 - 65 wurde Kategorie B vergeben, für Prozentrang 64 - 20 Kategorie C und die Kategorie D schließlich für Prozentrang 19-0. Dieses Modell ließ sich rechnerisch leicht umsetzen und transparent machen. Es bildete die Grundlage für die unternehmensinterne abschließende Diskussion und Entscheidung. Dabei wurde nur zwei Teilnehmern (statt ursprünglich fünf) die Potenzialkategorie A zugestanden, dennoch wurden Teilnehmer, die aufgrund des statistischen Entscheidungsmodells mit A oder B bewertet worden waren, weitergefördert.

6.4.6 Feedbackprozess

Insgesamt dauerte die diagnostische Maßnahme zwei Jahre. Nach der Implementation der ersten Auswahlstufe wurde die wissenschaftliche Begleitung zur Evaluation herangezogen, und gab nach jeder Auswahlstufe, also viermal, Rückmeldung über die jeweiligen Evaluationsergebnisse und natürlich viele deskriptive Informationen über den Verlauf der Maßnahme. Auch die deskriptiven Informationen, zum Beispiel Teilnehmerzahlen oder die diagno-

stischen Ergebnisse verschiedener Geschäftsbereiche gehören in diesem Sinne zu einer Treatment-Evaluation, da sie ebenfalls Hinweise liefern, ob die Maßnahme so umgesetzt wurde, wie geplant. Diese Informationen sind jedoch als solche für eine wissenschaftliche Dokumentation wie diese hier nicht so relevant, umso mehr, als sie ziemlich genau den Erwartungen entsprachen.

Auf diese Weise hatte die Evaluationsaktivitäten für die Auftraggeber nicht nur die eigentliche Evaluationsfunktion, sondern auch Dokumentationsfunktion und daneben unternehmensintern Legitimationsfunktion. Im Laufe des Evaluationsprozesses wurde von beiden Seiten zunehmend die Steuerungsfunktion der Evaluationsergebnisse auf die Maßnahme aufgegriffen, wie im folgenden beschrieben wird.

Neben den deskriptiven Daten zur Dokumentation der Maßnahme fielen weitere Informationen an, die auch im Hinblick auf die Interpretation der diagnostischen Daten und die Gestaltung der Maßnahme zu diskutieren waren. Es war auch eine Aufgabe der wissenschaftlichen Begleitung, dies möglichst kompakt und konstruktiv rückzumelden. Nicht alle Probleme, worauf die Evaluation stieß, ließen sich noch im Laufe der Maßnahme beheben. So waren die Anforderungsdimensionen verbindlich festgelegt sowie die Praxis, diese mit allen Instrumenten gleichermaßen zu erfassen, trotz der ernüchternden Resultate der Multitrait-Multimethod-Validierung.

Jedoch wurden vor dem Hintergrund der Evaluationsergebnisse Änderungen in der Maßnahme vorgenommen, gleichzeitig musste vor und während der Erarbeitung der Evaluationskriterien immer wieder rückgekoppelt werden, welche Fragestellungen und Ergebnisse für den Auftraggeber tatsächlich interessant waren. Damit war die gesamte Evaluationsstudie durch eine doppelte Rückkopplung charakterisiert, wobei die summativen Ergebnisse auch formative Funktion hatten, und umgekehrt Veränderungen der Maßnahme auch die Bearbeitung neuer Evaluationsfragen ermöglichten.

Nach der ersten Auswahlstufe im ersten Jahr zeigte sich, dass die mit der Datenintegration beauftragte Beratung ihr Verfahren höher gewichtete als das der Konkurrenz. Deshalb wurde die Integration im zweiten Jahr der anderen Beratung übertragen, die nun einen Bias zu Gunsten ihres Verfahrens zeigte. Deshalb sollte die wissenschaftliche Begleitung ein neutrales Entscheidungsmodell erarbeiten und umsetzen, wie im vorherigen Abschnitt beschrieben. Nach der zweiten Auswahlstufe stellte sich heraus, dass die Verfahren erwartungskonträr mit dem Assessment-Center korrelierten. Vor diesem Hintergrund entschied sich das Unternehmen, berufsnähere Verfahren in die Diagnostik einzubinden. Deshalb wurde im zweiten Jahr die Vorgesetztenbeurteilung hinzugenommen, und die Verfahren konnten an einem Berufserfolgskriterium validiert werden.

Da sich die theoretisch angenommene faktorielle Struktur der MBTI-

Werte anhand der Subskalenwerte replizieren ließ, aber weder eine inhaltliche Begründung noch ein Konzept für den Einsatz des Verfahrens und die Verwendung der Werte gefunden werden konnte, wie zum Beispiel eine ideale Ausprägung für Führungskräfte, wurden die MBTI-Werte von der Auswertung ausgeschlossen. Immerhin wurden auf Nachfrage die Skalenwerte vom MBTI zur Verfügung gestellt.

Einige Befunde waren darauf zurückzuführen, dass die Anforderungsdimensionen und ihr Bezug zu den Instrumenten nicht klar spezifiziert waren. Auch nach Rücksprache mit den Stakeholdern konnte manches nicht klarer gefasst und operationalisiert werden. Für die laufende Maßnahme hatte das keine Konsequenzen mehr, dennoch konnte die Studie für diese Thematik sensibilisieren. Das zeigte sich daran, dass ein Mitarbeiter der Personalabteilung nach seinem Wechsel zu einem anderen Unternehmen wiederum die wissenschaftliche Begleitung engagierte, um ein fundiertes Kompetenzmodell für Führungskräfte zu entwickeln, was in Kapitel 5.3 dargestellt wurde.

6.5 Fazit

In diesem Kapitel wurde beschrieben, wie eine diagnostische Maßnahme in der Praxis von wissenschaftlicher Seite evaluiert wurde, welche Strategien sich mit dem Schwerpunkt auf einer Treatment-Evaluation entwickeln lassen, wenn nur wenig externe Evaluationskriterien verfügbar sind, und mit welchen theoretischen und methodischen Einschränkungen dies verbunden ist, denen jedoch der Evaluator Rechnung tragen muss.

Anders als bei Untersuchungen, auch in der Anwendungsforschung, die von Anfang an in der Hand von Wissenschaftlern liegen, bestand hier die Herausforderung darin, dass der Evaluator wenig Einfluss darauf hatte, welche Daten in welcher Form zur Verfügung gestellt wurden. So ist es sehr schade, dass die Evaluationsdaten bereits stark aggregiert waren, und diese Aggregation, wie Nachfragen ergaben, vermutlich anhand klinischer Urteilsbildung erfolgte. Solche Voraussetzungen machten dieses Projekt gleichzeitig interessant, da sie unmittelbar die Praxisbedingungen widerspiegeln und das Projekt eine hohe praktische Relevanz aufweist.

Kapitel 7

Impact-Evaluation

7.1 Übersicht

Eine Impact-Evaluation bildet den Schwerpunkt in diesem Kapitel neben einigen Aspekten der Treatment-Evaluation. Das Problemlösemodell aus Kapitel 4.2.3 diente als Grundlage für eine diagnostische Untersuchung an 68 Nachwuchsführungskräften und als Rahmenkonzept zur Ableitung der Untersuchungshypothesen. Von besonderem Interesse war ein computergestütztes komplexes Szenario, das als differenzialpsychologisches Messinstrument konstruiert wurde, um das Anforderungsmodell an Problemlöser umzusetzen. Weitere Maße aus dem beruflichen und dem Persönlichkeitsbereich ergänzten die Prädiktoren. Fragestellung war, wie weit die Problemlöseleistung sowie die anderen Prädiktoren beruflichen Erfolg vorhersagen können, es wurde also eine konkurrente Kriteriumsvalidierung vorgenommen. Als berufliches Leistungskriterium wurden die Ergebnisse einer Arbeitsprobe modelliert. Außerdem sollte die Beziehung zwischen Problemlöseskalen und Maßen aus dem interpersonellen und dem Persönlichkeitsbereich geklärt werden.

Diese Untersuchung liefert einen Beitrag zur Anwendungsforschung. Als solche befindet sie sich im Spannungsfeld zwischen dem Ziel, wissenschaftliche Erkenntnisse zu gewinnen, und den Prioritäten der Kooperationspartner aus der Praxis. Das wirkt sich darauf aus, wie das Design gestaltet wird, wie die Daten organisiert sind und welche Auswertungen möglich sowie sinnvoll interpretierbar sind. Deshalb berichtet der folgende Abschnitt die Situation der diagnostischen Untersuchung in einer Praxisumgebung, bevor die Methoden und Ergebnisse im Zusammenhang mit der wissenschaftlichen Fragestellung dargestellt werden, die im Vordergrund steht. Auch für den Praxiseinsatz wird ein Fazit gezogen, wie brauchbar sich das computergestützte Szenario erwiesen hat.

7.2 Hintergrund

Die eignungsdiagnostische Untersuchung fand im Vertrieb Deutschland eines internationalen Unternehmens der pharmazeutischen Industrie statt. Dieses Unternehmen beauftragte dazu eine unabhängige Personalberatungsfirma, die Verfahren durchzuführen, auszuwerten, die Ergebnisse in ein diagnostisches Urteilsmodell zu integrieren und ein ausführliches schriftliches Gutachten für jeden Teilnehmer zu erstellen. Dabei setzte die Beratungsfirma teilweise eigene Verfahren ein. Weitere Verfahren bezog sie von ihren wissenschaftlichen Kooperationspartnern, die außerdem bei der Zusammenstellung des Instrumentariums berieten und an der Durchführung, Auswertung und Begutachtung für die gesamte Maßnahme beteiligt waren.

7.2.1 Praxisziele und Teilnehmer

Um intern Nachwuchsführungskräfte auszuwählen, die für zukünftige Leitungspositionen im Vertrieb geeignet waren, rief das pharmazeutische Unternehmen ein Programm ins Leben, das neben der Auswahl weitere Ziele aus dem Bereich der Führungskräfteentwicklung beinhaltete. Die Teilnahme an dem Programm war freiwillig, damit sich die Zielgruppe nur aus solchen Personen zusammensetzte, die sich für eine berufliche Weiterentwicklung interessierten. Auf der Basis der verwendeten eignungsdiagnostischen Verfahren wurden geeignete Teilnehmer für die Weiterförderung ausgewählt. Außerdem informierten die Ergebnisse jeden Teilnehmer über die Passung seiner Kompetenzen, Motivationen und Einstellungen mit den Anforderungen einer Position im mittleren Management im Vertrieb. Dies erfolgte in Form eines detaillierten Gutachtens sowie eines Feedbackgesprächs durch den Leiter der Personalentwicklungsabteilung.

Ein Ziel bestand darin, vorhandene Potenziale aufzudecken, um diese weiterfördern zu können. Daneben sollte das Programm den Teilnehmern helfen, im Zusammenhang mit den diagnostischen Verfahren die Ausprägung der eigenen Potenziale zu bewerten. Weiterhin sollte es ihnen zeigen, welche weiteren Entwicklungsmöglichkeiten neben der Vertriebsleitung denkbar waren. Außerdem sollten die Teilnehmer auf ihre problematischen Kompetenzbereiche hingewiesen werden, die sie mit Entwicklungsmaßnahmen verbessern konnten. Damit verbindet das Programm Zielsetzungen der Personalauswahl sowie der Personalentwicklung. Nach Pawliks Klassifikation diagnostischer Zielsetzungen (1976) sollte auf diese Weise Personenselektion mit Personenmodifikation kombiniert werden.

Das Programm zur Führungskräfteentwicklung stellte innerhalb der Personenselektion die erste Auswahlstufe für den Einstieg ins mittlere Mana-

gement des Unternehmens dar. Ihm folgte ein Assessment-Center für die Teilnehmer, die in der ersten Stufe als vielversprechend in die engere Wahl gezogen wurden. Die erfolgreichen Teilnehmer der zweiten Stufe sollten anschließend verbindlich darauf vorbereitet werden, Führungsfunktionen zu übernehmen.

7.2.2 Betriebliches Anforderungsmodell

Wie im Abschnitt 3.4.5 berichtet, greifen Praktiker zur Anforderungsanalyse häufig auf unstandardisierte, qualitative Methoden zurück, wie Einschätzungen von Experten, um Anforderungen an Führungskräfte zu erhalten. Auch hier wurde im Rahmen eines Workshops mit Metaplantechnik ein Anforderungsmodell für Führungskräfte im Vertriebsbereich formuliert. Teilnehmer des Workshops waren die Beratungsfirma sowie Vertreter des Unternehmens. Daraus resultierten acht so genannte Anforderungsdimensionen für eine Führungsposition, welche hier kurz so beschrieben werden, wie sie sich im Verständnis der Projektpartner darstellen. Nach diesen Anforderungen wurden die Daten aggregiert, was teilweise bei der wissenschaftlichen Auswertung zu berücksichtigen ist.

1. **Problemlösekompetenz** Der Teilnehmer kann in größeren Zusammenhängen denken, komplexe Sachverhalte schnell erfassen, strukturieren und gewichten. Dabei beherrscht er die Regeln des induktiven und deduktiven Denkens. Mehrere beeinflussende Faktoren berücksichtigt er zugleich. Er ist in der Lage, Problemlösungsalternativen zu entwickeln, zu fördern und zu bewerten, kann dabei mit Widersprüchen umgehen. Neue Ideen werden entwickelt, auf Realisierungsmöglichkeiten geprüft, bewertet und umgesetzt. Eine angemessene Risikobereitschaft ist vorhanden. Andere werden bei Bedarf mit Information versorgt.
2. **Planung, Koordination, Kontrolle** Im Idealfall verschafft sich der Teilnehmer einen Überblick und ist in der Lage, Prioritäten zu setzen. In seiner Arbeit geht er ökonomisch, ergebnisorientiert sowie pragmatisch vor und stimmt verschiedene Arbeitsabläufe aufeinander ab. Er ist in der Lage, Prozesse und Strukturen zu erkennen und mit deren Rahmenbedingungen effizient umzugehen. Die eigenen Arbeitsergebnisse will er kontinuierlich verbessern, kann bei Unübersichtlichkeit planvoll arbeiten und verarbeitet kurzfristige Veränderungen. Für sich und andere formuliert er realistische Arbeitsziele und Pläne. In Stresssituationen reagiert er angemessen und behält die Ruhe.
3. **Entscheidungsverhalten** Mit neuen Situationen kann der Teilnehmer umgehen. Die Entscheidung wird verfolgt und beibehalten, solange kei-

ne neuen Informationen eine Korrektur verlangen. Er versucht nicht, Entscheidungen hinauszuschieben. Er ist in der Lage, unter wirtschaftlichen oder organisatorischen Gesichtspunkten richtige Entscheidungen durchzusetzen und bedenkt die Folgen für sich und andere. Dabei geht er kalkulierbare Risiken ein. Um Dinge voranzutreiben, nutzt oder überschreitet er den Entscheidungsspielraum und stimmt sich nötigenfalls mit den Gruppenmitgliedern ab.

4. **Kommunikation** Der Teilnehmer geht von sich aus auf andere zu, ist höflich und zuvorkommend. Er tritt selbstbewusst auf, wirkt sympathisch und schafft Vertrauen. Er kann Kontakte schnell aufbauen und längerfristig pflegen. In Gesprächen und Verhandlungen drückt er sich klar und verständlich aus, kann diese zielgerecht steuern und stellt sich auf unterschiedliche Gesprächspartner ein. Argumente anderer kann der Teilnehmer zu eigenem Nutzen verarbeiten, gegensätzliche Standpunkte bringen ihn nicht aus dem Gleichgewicht.
5. **Integration, Kooperation, Konfliktbewältigung** In Gruppensituationen fühlt sich der Teilnehmer wohl. Er respektiert die Arbeit und die Meinung anderer. Bei der Zusammenarbeit liefert er beratende, aktive Beiträge und kennt die Voraussetzungen für gute Zusammenarbeit. Mit Konflikten geht er konstruktiv um, ohne dabei mehr als nötig von seiner Position abzuweichen.
6. **Persönliche Kompetenz** Der Teilnehmer identifiziert sich mit den Aufgaben des Unternehmens, ist bereit, Überstunden zu machen und sich außerordentlich zu engagieren, wenn es zum Wohle des Unternehmens ist. Er erträgt Ambivalenzen, kann mit Unsicherheiten anderer umgehen und sich bei Bedarf abgrenzen. Bei berechtigter Kritik ist er bereit, sein Verhalten zu überdenken und gegebenenfalls zu ändern. Misserfolge können in kurzer Zeit verarbeitet werden. Er ist dabei relativ unempfindlich gegen schlechte Laune anderer Menschen. Der Teilnehmer geht davon aus, dass er erfolgreich sein wird. Ihm wird die Führungsrolle angeboten, weil andere seine Kompetenz schätzen.
7. **Führungsverhalten** Anderen Menschen vertraut der Teilnehmer und gibt ihnen Selbstvertrauen. Er nimmt sich Zeit für seine Mitarbeiter und kann sie zur Erreichung gemeinsamer Ziele motivieren. Wegen seiner Ausstrahlung wird er in einer Führungsrolle akzeptiert und ist Ansprechpartner in der Abteilung. Aufgaben bearbeitet er mit überdurchschnittlichem Engagement. Der Teilnehmer kann delegierbare von nicht delegierbaren Aufgaben unterscheiden, kontrolliert die Zielerreichung und gibt Feedback. Gegenüber Mitarbeitern übt er sachliche,

konkrete und unmittelbare Kritik. Er vergibt geeignete Aufgaben, will Entwicklungsmöglichkeiten seiner Mitarbeiter entdecken und mit diesen umsetzen. Chancen für das Unternehmen erkennt und nutzt er. Seine Planung ist effektiv und ergebnisorientiert.

8. **Customer Focus** In seinem Handeln berücksichtigt der Teilnehmer Kundeninteressen, dabei priorisiert er die unternehmerischen Zielvorgaben. Er entwickelt auf den Kunden abgestimmte Verhandlungsstrategien. Dabei gelingt es ihm, flexibel auf Reaktionen des Verhandlungspartners einzugehen. Der Partner soll anhand schlüssiger Argumente überzeugt und nicht überredet werden. Schwierige Kundenprobleme löst der Teilnehmer, wobei er aktiv am Lösungsprozess mitwirkt. Sein Vorgehen orientiert er an der Marktrealität und erkennt, wann er bewährte Strategien einsetzen muss und wann Innovationen nötig sind. Er berücksichtigt das Zusammenspiel externer wirtschaftlicher Faktoren mit der Situation des Unternehmens. Betriebswirtschaftliche Daten werden adäquat erfasst und analysiert.

Für jede Dimension des Anforderungsprofils wurden verschiedene Verfahren herangezogen. Die Verfahren wurden zunächst einzeln ausgewertet und die Teilergebnisse auf die Anforderungsmerkmale bezogen. Ein Verfahren sollte Auskunft über die Kompetenzen eines Teilnehmers hinsichtlich mehrerer Dimensionen geben. Die optimale Ausprägung der Anforderungsdimensionen und damit der Erwartungshorizont wurde an einer Referenzstichprobe erfolgreicher Führungskräfte sowie aus weiteren Plausibilitätsüberlegungen abgeleitet. Die Einzelleistungen wurden anhand von Ratingskalen eingeschätzt und zu einem Gesamtwert verrechnet, für den ein Cutoff definiert wurde.

7.3 Durchführung

Im Rahmen des Programms zur Führungskräfteentwicklung wurden sechs Verfahren durchgeführt. Die Beratungsfirma stellte einen selbstentwickelten Fragebogen zur Selbstreflexion persönlicher Ziele und Interessen zur Verfügung. Weiterhin wurden zwei Persönlichkeitstests eingesetzt, sowie das Computerplanspiel FSYS-Forst. Zusätzlich ließ jeder Teilnehmer anhand eines weiteren Fragebogens einige Kompetenzen durch seinen Vorgesetzten einschätzen, ebenfalls ein eigenes Instrument der Beratungsfirma. Als Arbeitsprobe sollte die Teilnehmer eine schriftliche Präsentation zu einer fachbezogenen Fragestellung einreichen. Insgesamt nahmen 68 Personen an den eignungsdiagnostischen Untersuchungen zur Führungskräfteentwicklung teil,

40 Männer und 28 Frauen. Alle Teilnehmer hatten einen Hochschul- oder Fachhochschulabschluss, 11 Teilnehmer waren promoviert.

7.3.1 Computergestütztes Szenario

Das computergestützte Szenario FSYS (Wagener, 1994; Wagener, 2001) soll die Fähigkeit zum Lösen komplexer Probleme erfassen. Es wurde im Unterschied zu anderen deutschsprachigen Szenarios eigens als differenzialpsychologisches Messinstrument entwickelt und beruht mit seinen Skalen auf der in Abschnitt 4.2.3 bereits dargestellten hierarchischen Modifikation des Anforderungsmodells an Problemlöser (Dörner, 1986). Es hat eine ökologische inhaltliche Einbettung, die so umgesetzt wurde, dass besonderes Vorwissen in diesem Bereich keine Rolle für die Steuerleistung spielt. Bei FSYS muss der Teilnehmer einen simulierten forstwirtschaftlichen Betrieb zum wirtschaftlichen Erfolg führen. In fünf strukturidentischen Waldstücken pflanzt der Teilnehmer Bäume, zieht sie auf und holzt sie ab. Dabei muss er sich gegen Schädlinge wehren und auf die Qualität des Bodens achten.

Die Bearbeitungsdauer des Programms bleibt deutlich unter zwei Stunden. Ein interaktives Lernprogramm führt zunächst in die Programmbedienung ein, der eigentliche Testlauf dauert höchstens 90 Minuten. In jedem der 38 Szenario-Takte, die fiktive Monate darstellen, lassen sich Informationen über das simulierte System und die Maßnahmen einholen, Entscheidungen treffen und widerrufen. Sobald der Proband in den nächsten Takt geht, werden die jeweils getroffenen Entscheidungen umgesetzt. Eine Protokolldatei erfasst fortlaufend alle Bedienungsaktionen sowie die Systementwicklung. Sie wird vollautomatisch ausgewertet, sobald alle Takte durchlaufen wurden.

Das Auswertungsprogramm gibt die Steuerungsleistung in Form des Gesamtkapitals aus sowie weitere Skalen, denen das Anforderungsmodell von Dörner (1986) zugrunde liegt. Sie lassen sich den Bereichen Informationsmanagement, Selbstmanagement und Maßnahmenplanung zuordnen. Nicht abgedeckt sind Aspekte der Zielausarbeitung, weil das Steuerungsziel bereits durch die Instruktion vorgegeben ist. Die inhaltliche Zuordnung der Skalen ist durch ihren Anfangsbuchstaben gekennzeichnet. S-Skalen beziehen sich auf die Steuerleistung, was den Lösungsgütekriterien der meisten Szenarios entspricht. M-Skalen beinhalten Kriterien der Maßnahmenplanung, I-Skalen stehen für Informationsmanagement, E-Skalen für Selbstmanagement.

Als Indikatoren für die gesamte Steuerleistung dienen die Skalen SKAP und SKAPKOR. SKAP ist das erwirtschaftete Ergebnis im Sinne der Aufgabenstellung, bei SKAPKOR wurde das Ergebnis um die Auswirkung einzelner grober Steuerungsfehler teilweise kompensiert. ESICHER gibt an, wie häufig Entscheidungen pro Takt korrigiert oder widerrufen werden. ENOIRR

FSYS-Skala	Kurzbeschreibung
Steuerleistung	
SKAP	Entwicklung Gesamtvermögen
SKAPKOR	Grobfehlerkorrigiertes Gesamtvermögen
Maßnahmenplanung	
MNOFEHL	Vermeidung direkter Fehlsteuerung
MPRIORI	Angemessene Prioritäten
MRISIKO	Risikobereitschaft
MEFFIZI	Wirksamkeit von Maßnahmen
MVERSTA	Frühzeitiges Verständnis Gesamtsystem
Informationsmanagement	
IORIENT	Frühe Orientierung im System
IVORHAND	Informationsbeschaffung vor Maßnahmen
IFEED	Feedback einholen nach Maßnahmen
IFEEDG	Grafisches Feedback nach Maßnahmen
IKONTI	Kontinuierliche Kontrolle wichtiger Variablen
IKONTIG	Grafische Kontrolle wichtiger Variablen
Selbstmanagement	
ESICHER	Entschiedenheit bei Maßnahmen
ENOIRR	Vermeidung irrelevanter Handlungen
EWERT	Orientierung am Gesamtvermögen
EWERTG	Grafischer Abruf Gesamtvermögen

Tabelle 7.1: Übersicht über die FSYS-Skalen

beinhaltet, ob ein Proband erkennbar irrelevante Handlungen unterlässt. Wie häufig sich Probanden Tabellen oder Verlaufsgrafiken über das Gesamtvermögen des Betriebs anzeigen lassen, enthalten EWERT und EWERTG.

IORIENT, ein Indikator der frühen Orientierung im System, gibt an, wie weit sich ein Proband zu Beginn über die Zusammenhänge und den Zustand des Systems informiert. IVORHAND beinhaltet das Ausmaß, in dem sich der Proband über die Wirkungen von Maßnahmen erkundigt, bevor er sie anordnet. IFEED und IFEEDG sind Indikatoren dafür, wie häufig Ergebnisse von Maßnahmen numerisch oder grafisch abgerufen werden. IKONTI und IKONTIG stellen die Regelmäßigkeit dar, mit der wichtige Variablen überprüft werden.

Ob anhand des Informationsmenüs vermeidbare Fehleingriffe tatsächlich unterlassen wurden, resultiert aus MNOFEHL. Die Bereitschaft, hohe Risiken zwecks Gewinnsteigerung einzugehen, drückt sich in MRISIKO aus, typisch wäre dafür beispielsweise die Anpflanzung gewinnbringender Bäume, die besonders schädlingsanfällig sind. MEFFIZI beurteilt die Wirksamkeit

von Maßnahmen. Die Ausprägung von MVERSTA zeigt, wie stark der Proband bereits zu Beginn seine Handlungsmöglichkeiten ausschöpft. MPRIORI zeigt, wie sehr die Maßnahmen nach ihrer Dringlichkeit angeordnet werden. Dazu ermittelt eine künstliche Intelligenz für jeden Takt eine Prioritätenrangliste, mit der die gewählte Maßnahme verglichen wird.

Vor dem ersten Einsatz in einer realen Praxissituation hatte FSYS bereits eine dreijährige Entwicklungsphase an über 250 Versuchspersonen hinter sich, bei der zunächst die Oberfläche, Benutzerfreundlichkeit sowie Schwierigkeit optimiert wurde. Im Vergleich zu anderen Szenarios hat FSYS ein gehobenes Schwierigkeitsniveau, da es speziell für die Managementdiagnostik entwickelt wurde. Zudem wurde die Reliabilität ermittelt, zum einen durch Simulationsstudien, zum anderen durch Test-Retest-Korrelationen, wobei in der Wiederholungsuntersuchung ein strukturidentisches Szenario mit anderer semantischer Einbettung zum Einsatz kam (Wagener, 2001). Ebenso war bereits erfolgreich überprüft worden, ob die Beziehung der Skalen den erwarteten Modellzusammenhängen entsprachen. Außerdem wurden Indikatoren der Konstruktvalidität ermittelt.

7.3.2 Persönlichkeitstests

Neben dem Szenario FSYS kamen zwei standardisierte Persönlichkeitstests zum Einsatz, der mehrdimensionale Persönlichkeitstest für Erwachsene sowie der 16-PF in einer raschskalierten Version (Conrad, Mohr & Seydel, 1980). Der MPT-E (Schmidt, 1981) wurde speziell für die Berufs- und Betriebspraxis entwickelt. Er soll in erster Linie dazu dienen, Verhaltens- und Persönlichkeitsmerkmale zu erfassen, die für die berufliche Ausbildung und Tätigkeit relevant sind. Ein Konstruktionsziel war außerdem, dass das Verfahren solchen Personen zumutbar ist, für welche die Testsituation Prüfungscharakter hat. Beim MPT-E handelt es sich um einen Persönlichkeits-Struktur-Test. Er soll Aufschluss über charakteristische Verhaltensweisen von Personen geben und damit Rückschlüsse auf die berufliche Eignung der Probanden erlauben. Der mehrdimensionale Persönlichkeitsfragebogen für Erwachsene enthält 104 Items, die sieben Skalen zugeordnet werden können:

IS Ich-Schwäche

SE Soziale Erwünschtheit

RB Risikobereitschaft

RG Rigidität

AS Antriebsspannung

SZ Soziale Zurückhaltung

K Kontrollskala

Den MPT-E (Schmidt, 1981) bearbeiteten die Teilnehmer schriftlich, sie konnten ihn dazu nach Hause nehmen. Üblicherweise dauert die Bearbeitung etwa 30 Minuten. Es stehen für jedes Item vier Antwortmöglichkeiten zur Verfügung, die jeweils zweifach abgestufte Zustimmung oder Ablehnung ermöglichen. Ausgewertet werden die Antworten nur nach Zustimmung und Ablehnung. Dies soll den Probanden die Beantwortung der Items erleichtern, da sich die Probanden bei einer Frage durch eine undifferenzierte Antwort für ein Extrem entscheiden müssen. Das Verfahren kann als Einzel- oder Gruppentest durchgeführt werden. Die Objektivität ist durch die standardisierte Durchführung gewährleistet, jedoch haben einige Skalen Reliabilitätskennwerte unterhalb $r=0.60$. Als gut bis zufriedenstellend können dabei nur die Skalen SE, RG und AS gelten.

Beim 16-PF-Rasch (Conrad, Mohr & Seydel, 1980) handelt es sich um eine Überarbeitung der deutschen Version des Sixteen Personality Factor Questionnaire (16 PF). Auf Grundlage einer großen Gesamtstichprobe, hauptsächlich im Rahmen medizinisch- psychologischer Untersuchungen erhoben, wurde eine Analyse dieses Persönlichkeitsfragebogens mittels Rasch-Skalierung und Faktorenanalyse vorgenommen. Es zeigte sich, dass nur 82 der ursprünglichen 128 Items als homogene Skalenitems beibehalten werden konnten. Die von Cattell angenommenen 15 Persönlichkeitsfaktoren ließen sich nicht bestätigen. Stattdessen ergab sich eine siebenfaktorielle Lösung, wovon sechs Faktoren interpretierbar waren und die Testskalen der überarbeiteten Form darstellen.

A Introversion

B Emotionale Belastbarkeit

C Feinfühligkeit

D Anpassungsbereitschaft

E Antriebsstärke

G Nachdenklichkeit

Auch diesen Test bearbeiteten die Teilnehmer schriftlich außerhalb der Arbeitszeit auf einem separaten Antwortbogen mit vierfach abgestuften Ratingskalen. Sie hatten die Möglichkeit, ihn mit nach Hause zu nehmen. Seine Bearbeitung dauert etwa 20 Minuten. Von beiden Tests wurden zur Auswertung die Rohwerte herangezogen.

7.3.3 Selbst- und Fremdbeurteilung

Der Fragebogen zur *Selbstreflexion beruflicher Einstellungen, Ziele und Interessen* wurde von der Personalberatung entwickelt und zur Verfügung gestellt. Die Kandidaten bearbeiteten 31 offene Fragen schriftlich im eigenen Heim. Bei der Beantwortung der Fragen soll der Teilnehmer dazu angeleitet werden, möglichst konkret berufliche Vorstellungen zu präzisieren. Er bezieht sich deshalb auf die eigenen Motive, Fähigkeiten und Erfahrungen der Teilnehmer. Die Antworten sollen Rückschlüsse zulassen auf die Ausprägung kommunikativer und sozialer Fähigkeiten, persönlicher und fachlicher Kompetenz, auf das Führungspotenzial sowie auf das persönliche Engagement und den Ehrgeiz. Insgesamt gliedert sich der Fragebogen in vier thematische Bereiche auf, die nach Abschnitten geordnet sind.

Abschnitt A Ziele, Einstellungen und Werte bezüglich des Berufes

Abschnitt B Aufgaben im derzeitigen Tätigkeitsfeld

Abschnitt C Eigene Fertigkeiten und Zukunftserwartungen

Abschnitt D Führungskompetenz und Führungsmotivation

Aufgrund einer Referenzstichprobe erfolgreicher Führungskräfte im Vertrieb und theoretischen Erwartungen wurde ein Anforderungshorizont ausgearbeitet, der eine Bewertung der Antworten nach einem siebenfach abgestuften Ratingschema ermöglichte. Für jede der Ausprägungen von 7 (hervorragend) bis 1 (unzureichend) wurden typische Antworten und Mindeststandards beispielhaft formuliert, ähnlich wie verhaltensbezogene Ratingskalen beim Assessment-Center. Diese Ratingwerte ließen sich nach drei für die Auftraggeber interessanten Anforderungsbereichen zusammenfassen und dienten damit als Indikatoren für Integration, persönliche Kompetenz und Führungsverhalten.

Jeweils der direkte Vorgesetzte des Kandidaten füllte einen ebenfalls intern von der Personalberatung entwickelten Fragebogen zur *Fremdeinschätzung* aus. Er enthielt fünf offene Fragen, anhand derer der Vorgesetzte das Verhalten seines Mitarbeiters in führungsrelevanten Situationen sowie dessen Stärken und Schwächen frei beschreiben sollte. 25 weitere gebundene Items ermöglichten auf einer dreifach abgestuften Ratingskala eine Einschätzung der Kommunikationsfähigkeit, der Integration und Kooperation, der persönlichen Kompetenz, der Konfliktfähigkeit und des Führungsverhaltens.

7.3.4 Arbeitsprobe

Als eine Kostprobe ihrer fachlichen Kompetenz im Vertriebsbereich lieferten die Teilnehmer eine schriftliche Ausarbeitung zu einem von drei relevanten

Themen ab. Sie hatten etwa vier Wochen Zeit, die so genannte *Hausaufgabe* fertig zu stellen. Die Themen waren als offene Fallstudie gestellt, das heißt, zur eigentlichen Fragestellung gehörte eine etwa halbseitige Erläuterung der Situation und der Vorgaben. Aufgabe war es nicht nur, eine überzeugende Präsentation zu entwickeln, sondern sich dazu nötige Informationen auch in Form von Zahlenmaterial zu beschaffen, zu selektieren und zu verwerten. Fiktives Ziel war es, mit den ausgearbeiteten Vorschlägen die Geschäftsleitung zur Umsetzung und entsprechenden Investitionen zu bewegen. Dabei war die besondere Perspektive und Situation des eigenen Unternehmens zu berücksichtigen. Zu diesem Zweck standen drei Problemstellungen zur Wahl:

1. Kundenreklamation im Zeitalter der PCs und Online-Systeme
2. Geriatriische Patienten - eine neue Zielgruppe?
3. Großkunden, Einkaufsorganisation, Jahresgespräche - Ihre Antwort?

Die Ausarbeitungen wurden von den Beurteilern nach einem festgelegten Auswertungsschema bewertet, wobei vier Anforderungsbereiche nach jeweils acht verschiedenen Aspekten beurteilt wurden. Bei den Anforderungsbereichen handelte es sich um komplexes Problemlösen, den Bereich Planung, Koordination, Kontrolle, um Entscheidungsverhalten und Customer Focus. In Abhängigkeit davon, wie gut die Ausarbeitung die vier Anforderungsbereiche anhand der Aspekte abdeckte, wurden jeweils Ratingwerte von 0 (nicht erfüllt) bis 7 (voll erfüllt) vergeben.

7.4 Ergebnisse

Im vorhergehenden Abschnitt wurden insgesamt sechs Verfahren dargestellt, deren Ergebnisse sich wiederum nach den acht betrieblichen Anforderungsdimensionen aufteilen lassen, die ebenfalls Unterbereiche besitzen. Zudem haben die meisten Verfahren eigene Skalenwerte, die sich zwar den Anforderungsdimensionen zuordnen lassen, aber außerdem eine eigene inhaltliche Bedeutung haben. Damit resultiert aus der Studie eine Fülle von korrelierbaren, aggregierbaren Daten und Auswertungsmöglichkeiten. Nicht nur im Interesse einer übersichtlichen Darstellung, sondern auch aus wissenschaftsstrategischen und statistischen Gründen muss die Auswertung auf wenige, theoriegeleitete Analysen beschränkt werden.

Deshalb werden die vorliegenden Daten nur zur Beantwortung weniger, im Kontext der vorliegenden Arbeit sinnvoller Fragen ausgewertet. Dabei handelt es sich um die Frage, wie weit Skalen zur Erfassung des komplexen

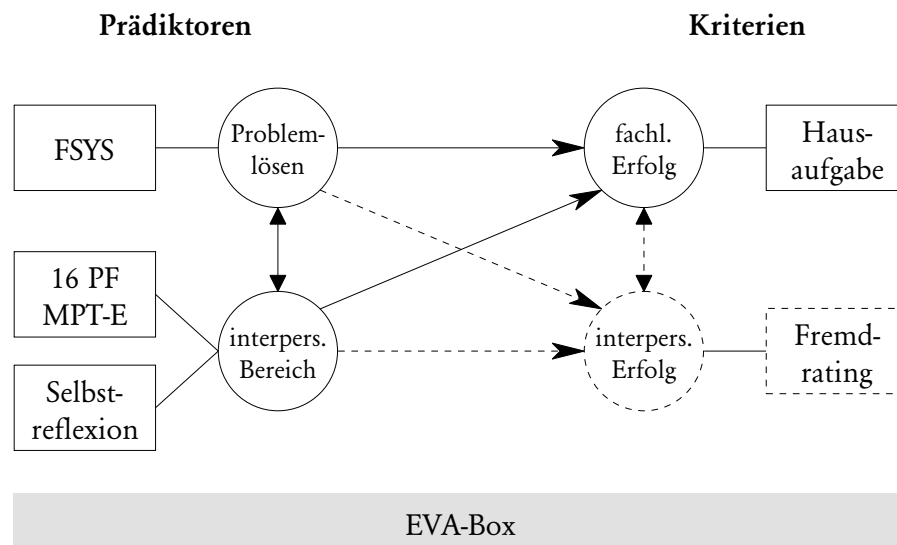


Abbildung 7.1: Zuordnung der Verfahren zu Konstrukten auf Prädiktor- und Kriteriumsseite.

Problemlösens neben anderen Prädiktoren zur Vorhersage beruflichen Erfolgs beitragen und wie das Verhältnis der Problemlösemaße zu den Maßen aus dem interpersonellen und dem Persönlichkeitsbereich aussieht, da ja bei der Literaturanalyse vermutet wurde, dass es Überschneidungsbereiche gibt. Eine weitere Einschränkung der Auswertungsmöglichkeiten ergibt sich durch die Probleme der Fremdeinschätzung, deren Ergebnisse nur eingeschränkte Varianz aufweisen.

7.4.1 Evaluationsstrategie

Die große Anzahl von Variablen legt nahe, komplexere multivariate Verfahren durchzuführen, die im Verhältnis dazu kleine Stichprobe von Teilnehmern lässt dagegen wenig zu, was über korrelative Analysen hinausgeht. Zudem können nicht alle denkbaren Analysen, zum Beispiel in Form von Korrelationen, gerechnet werden, da man sonst das Signifikanzniveau in einen sehr ungünstigen Bereich korrigieren müsste, um sich gegen die Wahrscheinlichkeit nur zufällig signifikanter Ergebnisse abzusichern. Deshalb sind zunächst Modelle für die Kriterien und die Prädiktoren zu entwickeln, um anschließend in diesem Sinne die Variablen theoriegeleitet zu aggregieren und auszuwerten. Dafür wurden folgende Untersuchungshypothesen aufgestellt, die kompakter in der Abbildung 7.1 visualisiert sind:

1. Die Prädiktoren, welche das komplexe Problemlösen erfassen, hängen stärker mit den problemlösebezogenen Kriterien zusammen als mit denen zur Erfassung interpersoneller Fertigkeiten und Persönlichkeitseigenschaften.
2. Die Prädiktoren, welche Persönlichkeitseigenschaften und interpersonelle Fertigkeiten erfassen, hängen stärker mit den persönlichkeitsbezogenen Kriterien zusammen als mit denen zur Erfassung der Problemlösefähigkeit.
3. Die beiden Prädiktorgruppen hängen untereinander weniger stark zusammen als jeweils mit dem korrespondierenden Kriterien.
4. Die beiden Kriteriengruppen hängen untereinander weniger stark zusammen als jeweils mit dem korrespondierenden Prädiktor.

Als Prädiktoren kommen generell die Items oder Skalen der Persönlichkeitstests, des Selbstreflexionsbogens sowie die FSYS-Skalen in Frage. Typischerweise lässt sich dazu die Beurteilung durch den Vorgesetzten als Kriterium heranziehen, außerdem die Arbeitsprobe in Form der Hausaufgabe. Weiterhin sollten sowohl das Kriterium als auch die Prädiktoren ein vergleichbares Aggregationsniveau (Murphy & Shiarella, 1997) und vergleichbaren Inhalt haben (Ashton, 1998), um eine sinnvolle Beziehung herzustellen, weshalb diese hier analog zu modellieren sind. Aus der Perspektive dieser Arbeit bestehen die wesentlichen Inhalte in den Anforderungen an Problemlöser sowie im interpersonellen Bereich.

Für das Problemlösen sind auf Prädiktorenmite die FSYS-Skalen heranzuziehen. Auf Kriterienseite bietet sich dazu die Hausaufgabe an, da im Zusammenhang mit ihr Strategien der Informationsverarbeitung, der Wissensorganisation und der Prioritätensetzung wichtig sind. Für den interpersonellen Bereich könnten als Kriterien die Items der Fremdeinschätzung dienen, als Prädiktoren Persönlichkeitsskalen und Aspekte des Selbstreflexionsbogens. Leider zeigte sich, dass die Ratings, mit welchen die Vorgesetzten ihre Mitarbeiter einschätzten, nur sehr geringe Varianz aufwiesen, da sie meistens den höchsten Wert vergaben, wahrscheinlich ein Milde-Effekt. Plausibel wird dies durch die Untersuchungssituation, da die Teilnehmer wussten, von wem sie eingeschätzt wurden, was wiederum den Vorgesetzten bekannt war.

Damit fällt dieses inhaltlich sehr interessante Kriterium durch seine methodischen Probleme aus, obwohl sich die Items mit ihren Inhalten den Teilkategorien des interpersonellen Bereichs aus dem Kategoriensystem zuordnen ließen. So entfällt die Überprüfung der Frage, welche Zusammenhänge zwischen dem problemlösebezogenen Kriterium und dem interpersonellen Kriterium bestehen und wie gut es durch die Problemlöseskalen und die Indikatoren für den interpersonellen Bereich im Vergleich vorhergesagt wird. Die

dadurch ausgeschiedenen Auswertungsmöglichkeiten sind in Abbildung 7.1 durch gestrichelte Linien angezeigt.

Übrig bleibt noch, zu prüfen, wie groß der Überschneidungsbereich zwischen den beiden Prädiktorengruppen ist und wie gut beide das problemlösebezogene Kriterium vorhersagen. Nach den erwähnten Überlegungen zur Beziehung von Kriterien und Prädiktoren ist anzunehmen, dass die Prädiktoren mit dem inhaltlich gleichen Bezug das Kriterium besser vorhersagen als die inhaltlich anders konzipierten Prädiktoren, das wären die Persönlichkeitsskalen und Skalen aus der Selbstreflexion. Im Mittelpunkt steht jedoch der Zusammenhang zwischen den Problemlösemaßen aus FSYS als Prädiktor und den problembezogenen Skalen der Arbeitsprobe.

7.4.2 Aggregation

Faktor (Varianzanteil)	STRUK (28%)	INFO (26%)	PRAG (16%)
Inhalte strukturieren	.87	.26	.16
Gesamtes erfassen	.82	.17	.17
Übersichtlichkeit	.78	.24	.09
Klare Fragestellung	.78	.29	.21
Wesentliches	.74	.16	.25
Prioritäten setzen	.66	.34	.32
Abstimmung	.14	.79	.41
Schlüssigkeit	.38	.76	.16
Adäquate Quellen	.49	.76	.01
Versch. Quellen	.24	.75	-.15
Praxisbezug	-.03	.75	.45
Ziele - Maßnahmen	.25	.75	.29
Umsetzungsvorschläge	.28	.12	.85
Arbeitsprozesse	.23	.13	.80
Hinterfragen	.45	.41	.57
Bewertung	.24	.50	.11

Tabelle 7.2: Faktorstruktur der Ratingwerte für die Arbeitsprobe nach Varimax-Rotation (N=60).

Dazu waren die Ratings der Hausaufgabe entsprechend zu aggregieren. Diese hatten durch die Vorgaben des betrieblichen Anforderungsmodells bereits eine etwas andere Ausrichtung erhalten, als dies inhaltlich in Bezug auf das Problemlösemodell und die FSYS-Skalen sinnvoll wäre. Aggregierte man die Werte nach inhaltlichen Aspekten, so dass sie sich auf die Problemlöseskalen beziehen ließen, waren sie stark interkorreliert, was wahrscheinlich auf Halo-Effekte zurückzuführen ist, da die Rater ungeschult waren. Deshalb

wurde eine Faktorenanalyse durchgeführt. Die nach Kaiser-Kriterium extrahierten drei Faktoren wurden nach einer Varimax-Rotation interpretiert.

Nach ihrem Korrelationsmuster mit den Hausaufgabenratings wurden die drei Faktoren als Pragmatismus (PRAG), Informationsmanagement (INFO) und Strukturiertheit (STRUK) bezeichnet, die einzelnen Faktorladungen sind in Tabelle 7.2 zu sehen. Pragmatismus beinhaltet unter der Perspektive des betrieblichen Anforderungsmodells vor allem, ob in der Hausaufgabe realisierbare Ideen enthalten waren, ob konkrete Vorschläge zur Umsetzung gemacht wurden und inwieweit diese die Defizite alter Konzepte explizit berücksichtigen. Informationsmanagement bezieht sich vor allem darauf, ob der Teilnehmer adäquate Informationsquellen nutzte, logisch, schlüssig und differenziert argumentierte und die Abläufe in seinem Konzept aufeinander abstimmte. Mit Strukturierung ist schließlich gemeint, ob die Fragestellung und damit verbundene Probleme vollständig, aber aufs Wesentliche konzentriert erfasst, die Darstellung inhaltlich strukturiert erfolgt und optisch entsprechend veranschaulicht wird.

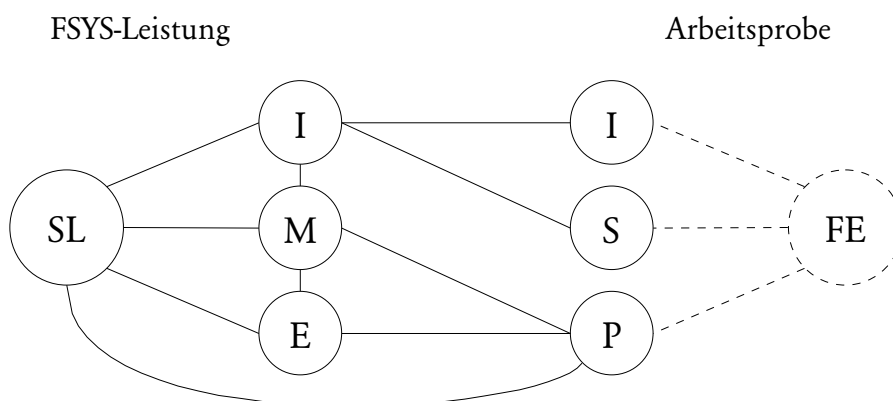


Abbildung 7.2: Operationalisierte Problemlösemaße auf Prädiktor- und Kriteriumsseite sowie erwartete Zusammenhänge.

Deshalb sollte der Pragmatismusfaktor (P) vor allem mit dem Gesamtergebnis und den M-Skalen zusammenhängen, da diese am unmittelbarsten auf das Ergebnis bezogen sind sowie einigen E-Skalen. Das Informationsmanagement (I) der Hausaufgabe sollte vor allem mit den Skalen des Informationsmanagements zusammenhängen, die Planungs- Überwachungsaktionen beinhalten, während von der Strukturierung (S) Beziehungen zum Umgang mit bereits aufbereiteter Information erwartet werden. Die drei Faktoren sind als Aspekte des allgemeinen Kriteriums fachlicher Erfolg (FE) gedacht. Dieses Beziehungsmuster ist in Abbildung 7.2 veranschaulicht.

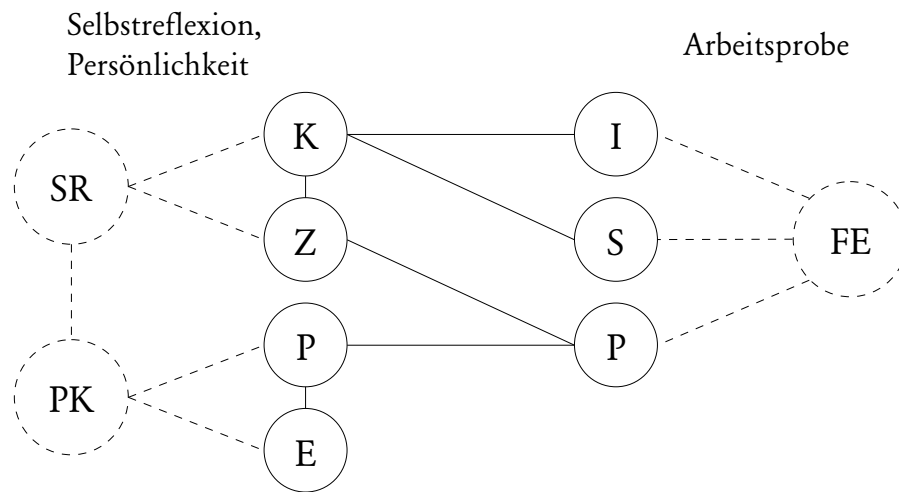


Abbildung 7.3: Aggregate aus Selbstreflexion und Persönlichkeit und erwartete Zusammenhänge mit den Faktoren der Arbeitsprobe.

Für die Zusammenhänge zwischen den FSYS-Skalen und den anderen beiden Prädiktoren, den Persönlichkeitsskalen und den Fragen zur Selbstreflexion, musste ebenfalls zunächst eine theoretische Erwartung formuliert werden, um die Daten entsprechend zu aggregieren. Aus den Persönlichkeitsskalen wurden zwei Aggregate gebildet. Engagement (ENG) im Sinne eines aktiven, eventuell unvorsichtigen Herangehens setzt sich zusammen aus den MPT-E-Skalen Flexibilität, Risikobereitschaft und Ich-Stärke sowie den 16-PF-Skalen Antriebsstärke, Belastbarkeit und Anpassungsbereitschaft. Pragmatismus (PRAG) überschneidet sich inhaltlich teilweise mit Engagement, bedeutet jedoch eher ein überlegtes, zielstrebiges und ergebnisorientiertes Vorgehen. Es beinhaltet die MPT-E-Skalen Flexibilität und soziale Erwünschtheit sowie die 16-PF-Skalen Antriebsstärke und schwächer gewichtet, Risikobereitschaft, Anpassungsbereitschaft und soziale Erwünschtheit.

Aus dem Fragebogen zur Selbstreflexion wurden dagegen zwei überschneidungsfreie Aggregate aus der ungewichteten Addition zweier Fragebogenkomplexe gebildet. Kompetenz (KOMP) bezieht sich auf verschiedene Fragen zur Einschätzung eigener beruflicher Kompetenzen und konkreten Vorstellungen, wie sich diese Kompetenzen in aktuellen Aufgaben und zukünftigen Positionen einbringen lassen. Zielstrebigkeit (ZIEL) wiederum beinhaltet Items, die nach persönlichen Berufszielen und deren Bewertung fragen.

Ewartet wird, dass Engagement als ein Verhaltensstil, energisch und rasch zu handeln, sich vor allem auf die Wirksamkeit von FSYS-Maßnahmen aus-

wirkt, wobei selten gezögert oder korrigiert wird. Dies sollte jedoch, da indirekt wirksam, weniger Einfluss auf die Gesamtsteuerleistung haben. Pragmatismus als ein mehr ergebnisorientierter Stil sollte stärker mit der Gesamtleistung und außerdem mit der Prioritätensetzung in Zusammenhang stehen. Ähnliches gilt für die Zielstrebigkeit. Kompetenz beinhaltet den Erwerb und gezielten Umgang mit dem eigenen Fachwissen, weshalb substanzielle Korrelationen mit Informationsmanagement, Prioritätensetzung und Steuerleistung zu erwarten sind.

Zuletzt bleibt noch die Frage, wie weit die Persönlichkeits- und Selbstreflexionsaggregate mit der Arbeitsprobe korrelieren. Da sie inhaltlich analog gebildet wurden, sollte die Zielorientierung (Z) vor allem mit dem Pragmatismus aus der Hausaufgabe korrelieren, die Kompetenz (K) mit den beiden informationsbezogenen Skalen Strukturierung und Informationsmanagement. Das Persönlichkeitsaggregat Pragmatismus (P) sollte entsprechend mit dem Pragmatismus der Hausaufgabe korrelieren. Engagement (E) mit den eher motivationalen Inhalten findet keine Entsprechung in den Faktoren der Hausaufgabe, insofern sollte es auch nicht bedeutsam korrelieren. Abbildung 7.3 stellt dieses Zusammenhangsmuster als Übersicht dar.

7.4.3 Kriteriumsvalidität Problemlösen

Im vorherigen Abschnitt wurden die Auswertungsmöglichkeiten durch mehrere Entscheidungen und Operationen erheblich eingeschränkt. So wurden zunächst in Anlehnung an das Problemlösemodell Prädiktoren und Kriterien so modelliert, dass sie inhaltlich und vom Aggregationsniveau her prinzipiell vergleichbar waren. Auf dieser Grundlage konnten die Indikatoren aus den einzelnen Verfahren identifiziert und verdichtet werden, um daraus eine wesentlich kleinere Menge an korrelierbaren Maßen zu gewinnen, für die außerdem Erwartungen an die korrelativen Zusammenhänge formuliert waren.

Wie man daraus weiterhin ableiten kann, sind für alle dargestellten Beziehungen positive korrelative Zusammenhänge zu erwarten, weshalb die Korrelationen einseitig auf einem Signifikanzniveau von $p < 0.05$ getestet werden. Den aggregierten Maßen sowie den FSYS-Skalen wird Intervallskalenniveau unterstellt, weshalb sich Pearson Produkt-Moment-Korrelationen bestimmen lassen. Um Scheinkorrelationen durch Ausreißer zu identifizieren, wurden außerdem Rangkorrelationen berechnet und mit den Produkt-Moment-Korrelationen verglichen, dies trat jedoch kaum auf. Ohnehin sollte die Stichprobe groß genug sein, dass einzelne Ausreißer keine wesentlichen Verzerrungen bewirken können. Nach fallweisem Ausschluss von Missings basieren die dargestellten Korrelationen auf einer Stichprobe von 58 Teilnehmern.

FSYS	Hausaufgabe			
	PRAG	INFO	STRUK	
SKAP	.32* (e)	-.08		-.14
SKAPKOR	.40* (e)	-.05		.21
MPRIORI	.32* (e)	.04		.19
MEFFIZI	.23* (e)	.01		.12
ESICHER	.34* (e)	-.11		.12
EWERT	.16 (e)	-.31*		.20
ORIENT	.11	.39* (e)		.08
IVORHAND	.11	.39* (e)		.07
IFEED	.28*	-.16 (e)		-.04
IKONTI	.16	-.18 (e)		.15
IFEEDG	.07	.12		.23* (e)
IKONTIG	.14	.12		.25* (e)

Tabelle 7.3: Korrelationen zwischen FSYS-Skalen und Arbeitsprobe, N=58, * signifikant ($p < .05$), (e) erwartete Zusammenhänge.

Es wurde jedoch darauf verzichtet, die Resultate gegen Zufallseffekte in den immer noch insgesamt 88 Einzelkorrelationen abzusichern, indem eine entsprechende Korrektur des Signifikanzniveaus vorgenommen würde. Zum einen gibt es noch keine Untersuchung dieser Art, und die Zielsetzung ist eher explorativ, da die Ergebnisse zur diagnostischen Modellbildung beitragen sollen. Damit wird in dieser Phase des Forschungsprozesses lieber in Kauf genommen, zu Unrecht Zusammenhänge zu unterstellen, als potenziell bestehende und wichtige zu übersehen. Zum anderen sind bereits Hypothesen dazu formuliert, an welchen Stellen signifikant positive Korrelationen erwartet werden und an welchen Stellen keine Zusammenhänge bestehen sollten. Die Beschränkung auf positive Korrelationen und die Interpretation vor dem Hintergrund gerichteter Hypothesen reduziert den Einfluss von Zufallseffekten noch weiter.

Tabelle 7.3 zeigt die Korrelationen der FSYS-Problemlösemaße mit den drei problemrelevanten Faktoren der Hausaufgabe, die als Arbeitsprobe ein Berufserfolgskriterium, vor allem für fachlichen Erfolg, darstellt. Erwartete Zusammenhänge sind mit (e) angezeigt. Wie zu vermuten war, korrelierten die Gesamtlösungsgütekriterien SKAP und SKAPKOR hoch mit Pragmatismus, das Gleiche gilt für MPRIORI, MEFFIZ und ESICHER, jedoch nicht für EWERT. Die signifikante Korrelation mit IFEED widerspricht der Erwartung.

Mit Informationsmanagement aus der Hausaufgabe korrelierten die Indikatoren planerische Informationsbeschaffung IVORHAND und IORIENT signifikant, die Überwachungsaktivitäten, erfasst als IFEED und IKONTI,

entgegen der Voraussage negativ, wenn auch nicht signifikant. Ein signifikanter negativer Zusammenhang mit EWERT war eine der wenigen durch Ausreißer verursachten Scheinkorrelationen. Grafischer Abruf von Informationen (IFEEDG, IKONTIG) korrelierte erwartungskonform mit der Strukturierung als letzter Hausaufgabenfaktor. Insgesamt lässt sich also festhalten, dass die FSYS-Skalen substantiell und erwartungsgemäß mit den korrespondierenden Faktoren der Hausaufgabe korrelieren.

FSYS	Persönlichkeit			Selbstreflexion			
	ENG		PRAG	KOMP		ZIEL	
SKAP	.13		.30* (e)	.18 (e)		.29* (e)	
SKAPKOR	.21*		.32* (e)	.19 (e)		.32* (e)	
MEFFIZI	.37* (e)		.24*	-.06		.04	
ESICHER	.21* (e)		.36* (e)	.21* (e)		.20	
MPRIORI	.10		.10 (e)	.31* (e)		.37* (e)	
EWERT	.00		.07	.06		.12 (e)	
IORIENT	.11		.06	.32* (e)		.21	
IVORHAND	.14		.10	.34* (e)		.18	
IFEEDG	.03		.04	.31* (e)		.07	
IKONTIG	-.05		-.04	.35* (e)		.14	

Tabelle 7.4: Korrelationen zwischen FSYS-Skalen und aggregierten Maßen der Persönlichkeits- und Selbstreflexionsfragebogen, N=58, * signifikant ($p < .05$), (e) erwartete Zusammenhänge.

7.4.4 Konstruktvalidität

In Tabelle 7.4 befinden sich die Korrelationen zwischen FSYS und den Persönlichkeitsaggregaten sowie den Skalen des Selbstreflexionsbogens. MEFFIZI, die Intensität von Maßnahmen im Szenario, und ESICHER, einmal angeordnete Maßnahmen ohne Zögern umzusetzen, korrelierten erwartungsgemäß mit dem Persönlichkeitsaggregat Engagement als schwungvollem, aktivem Herangehen. Mit Pragmatismus korrelierten ganz wie vermutet deutlich die Gesamtsteuerleistung und außerdem mit ESICHER, jedoch weniger die Prioritätensetzung (MPRIORI), unerwartet außerdem MEFFIZI.

Die FSYS-Skalen zum Informationsmanagement wiesen erwartungsgetreue Zusammenhänge mit der aus dem Selbstreflexionsbogen modellierten fachlichen Kompetenz auf, die außerdem mit Prioritätensetzung und zügigem Entscheidungsverhalten korrelierte. Die Korrelationen mit der Steuerleistung lagen unter den Erwartungen. Dagegen sagten diese Skalen wie vermutet die berufliche Zielstrebigkeit vorher, ebenso war der Zusammenhang mit der

Prioritätensetzung hoch, weniger mit der Wertorientierung (EWERT). Insgesamt ergab sich auch hier eine günstige Bilanz, da die überzähligen positiven Korrelationen im Persönlichkeitsbereich aus der Interkorrelation der beiden Maße resultierten. Nur der ausbleibende Zusammenhang zwischen Pragmatismus und MPRIORI ließ sich nicht auf diese Weise auflösen.

	Hausaufgabe					
	PRAG		INFO		STRUK	
KOMP	.40*	(e)	.41*	(e)	.20	
ZIEL	.20		.24*		.34*	(e)
PRAG	.03		.02		.14	(e)
ENG	-.01		-.01		.11	

Tabelle 7.5: Korrelationen zwischen Selbstreflexion, Persönlichkeit und Arbeitsprobe, N=58, * signifikant ($p < .05$), (e) erwartete Zusammenhänge.

Zuletzt bleibt noch die Frage übrig, was neben FSYS die anderen beiden Prädiktoren zur Vorhersage der beruflichen Leistung in Form der Arbeitsprobe leisten. Tabelle 7.5 zeigt die Korrelationen der beiden verdichteten Maße aus dem Selbstreflexionsbogen sowie der beiden Persönlichkeitsaggregate mit den drei Faktoren der Hausaufgabe. Dabei korrelierte die berufliche Zielstrebigkeit wie erwartet mit dem Pragmatismusfaktor, allerdings auch mit dem Informationsmanagement. Die Kompetenz in Form der gezielten Anwendung von Fachkenntnissen mit den beiden informationsbezogenen Skalen Strukturierung und Informationsmanagement. Da Engagement inhaltlich keine Entsprechung in den Hausaufgabenskalen findet, sollte es auch nicht bedeutsam mit ihnen korrelieren. Allerdings sollte das Persönlichkeitsmaß Pragmatismus mit dem analog konzipierten Hausaufgabenfaktor zusammenhängen, dies war hier nicht der Fall.

Hier wurden also drei von vier erwarteten Zusammenhängen gefunden, in den restlichen acht Vergleichen sollten keine Zusammenhänge bestehen, es wurde aber einer angetroffen. Das ist insgesamt zufriedenstellend, man sollte jedoch berücksichtigen, dass sowohl die Arbeitsprobe wie auch der Selbstreflexionsbogen von denselben Ratern ausgewertet wurde, also der hohe Zusammenhang teilweise durch einen Halo-Effekt erklärt werden könnte. Die durchgehend geringen Korrelationen mit den Persönlichkeitsskalen lassen sich möglicherweise darauf zurückführen, dass trotz inhaltlicher Übereinstimmung die Persönlichkeitsitems weniger spezifisch als die Hausaufgabenitems waren.

7.5 Fazit

Diese Untersuchung hatte als Schwerpunkt eine Impact-Evaluation als diagnostische Anwendung des Problemlösemodells, ergänzt um den interpersonellen Bereich. Selbst wenn man die methodischen Einschränkungen mit einbezieht, die sich in dieser Praxissituation ergaben, lässt sich diese Untersuchung durch die modellkonformen Korrelationsmuster als erfolgreich bewerten, wobei es etliche Verbesserungsmöglichkeiten gibt. So wäre es im Sinne der Fragestellung günstiger gewesen, wenn das diagnostische Instrumentarium sowie die Auswertung direkt und durchgängig auf der Basis des hier erarbeiteten Modells implementiert wäre, statt die Daten, die eigentlich auf dem betrieblichen Anforderungsmodell beruhen, entsprechend neu zu ordnen.

Außerdem wurden der Selbstreflexionsbogen und die Hausaufgabe durch Rater ohne psychologische Ausbildung ausgewertet. Zwar bekamen die Rater detaillierte Auswertungsrichtlinien mit Ankerbeispielen für die Rating-Stufen, jedoch war nicht festzustellen, wie weit sie sich daran hielten. Ein ähnliches Problem stellt die Fremdeinschätzung durch die Vorgesetzten dar, die davor zurückschreckten, die ganze Bandbreite der Ratings auszunutzen. Das ist nicht nur für die wissenschaftliche Auswertung problematisch, sondern beeinträchtigt ebenso den praktischen Nutzen dieser Daten. Im Falle der Selbstreflexion und der Hausaufgabe ließen sich Urteilsverzerrungen durch die Auswertungsstrategien teilweise kompensieren, das Vorgesetztenurteil war aufgrund der geringen Varianz unbrauchbar geworden.

Gerade diese Probleme, die durch verschiedene Urteilsfehler bei menschlichen Ratern entstehen können, hatte das computergestützte Szenario nicht und es war gleichzeitig wesentlich ökonomischer, vergleicht man beispielsweise den Zeitbedarf für die Beantwortung und Auswertung der offenen Fragen des Selbstreflexionsbogens. Es scheint also selbst im Kontext des außerwissenschaftlichen Einsatzes sehr robust, da es die Durchführungs- und Auswertungsobjektivität optimiert. Auch die Teilnehmer kamen gut damit zurecht, was an dem reibungslosen Ablauf zu sehen war. Lediglich die Interpretationsobjektivität könnte bezogen auf einzelne Skalen beeinträchtigt sein, wenn es den Untersuchungsleitern an Fachkenntnis mangelt. Die Skalen lieferten außerdem verwertbare und konkrete Anhaltspunkte zur Gutachtenerstellung.

Abgesehen vom computergestützten Szenario, das explizit auf der Grundlage des hier verwendeten Problemlösemodells entwickelt wurde, lassen sich für die weiteren hier verwendeten Verfahren Alternativen denken, insbesondere die Indikatoren für die berufliche Leistung sind ein weites Feld. Möglicherweise wäre auch eine Managementdiagnostik auf der Grundlage anderer Konstrukte erfolgreich. Dennoch lässt sich festhalten, dass das verwendete

Problemlösemodell als Hintergrund für die Aggregation der Daten und für die Evaluationsstrategie offensichtlich eine brauchbare Grundlage lieferte. Mit Korrelationen im Bereich von $0.30 < r < 0.40$ resultierten aus den errechneten Koeffizienten für die konkurrente Kriteriumsvalidität nach der Einteilung von Cohen (1988) mittlere Effektgrößen.

Kapitel 8

Outcome-Evaluation: Forschungsmodell

8.1 Übersicht

In den vorhergehenden Kapiteln wurde hauptsächlich aus einer Anwenderperspektive sowie induktiv-empirisch beleuchtet, auf welchen Anforderungsmodellen eine Managementdiagnostik beruhen könnte und welche Ziele eine normative Outcome-Evaluation (Chen, 1990) abdecken sollte. Gleichzeitig sollte die fachlich-wissenschaftliche Perspektive nicht vernachlässigt werden. Deshalb soll hier eine Strategie vorgestellt werden, wie sich die Ergebnisse der Managementforschung anhand eines psychologischen Anforderungsmodells systematisieren lassen.

Rosenstiel (1992) zufolge muss ein Eignungsdiagnostiker wissen, wofür eine Person geeignet sein soll. Deshalb sollte eine Anforderungsanalyse jeder Diagnostik vorausgehen (Kannheiser, 1995). Auch bei Führungsaufgaben sei demzufolge zu bestimmen, welche Anforderungen eine Tätigkeit enthält, um spezifische Personenmerkmale zu identifizieren. In der Literatur erfährt man wenig Konkretes darüber, wie man von den Tätigkeits- zu den Personenmerkmalen gelangt. Zudem lassen sich bei Führungskräften Tätigkeitselemente schwerer identifizieren und empirisch erfassen.

Schließlich ist ein diagnostisches Modell zu finden, dass sich mit den Aspekten einer Führungstätigkeit verbinden lässt. Dafür scheint Dörners Anforderungsmodell an Problemlöser (1986) geeignet zu sein, insbesondere um Informationen aus solcher Führungsliteratur systematisieren zu können, die sich auf ähnlichem Abstraktionsniveau befindet. Im Folgenden wird eine qualitative Methode zur literaturgestützten Anforderungsanalyse als konzeptueller Beitrag zur Outcome-Evaluation entwickelt und umgesetzt.

8.2 Methode

Als ein möglicher Zugang zu einer solchen Anforderungsanalyse lässt sich eine qualitative Literatursystematisierung betrachten. Diese greift auf Techniken zurück, die teilweise aus anderen Disziplinen oder anderen psychologischen Fragestellungen stammen und hier zu einer adäquaten Methode kombiniert werden. Damit soll ein metaanalytisches Vorgehen ergänzt werden, wenn man begrifflich heterogene, rein konzeptuelle oder interdisziplinäre Arbeiten zusammenfassen möchte und die typischen Probleme eines narrativen Reviews vermeiden will.

Aus diesen Gründen wurde als Technik zur Auswertung von sprachlichem Material die qualitative Inhaltsanalyse nach Mayring (1997) zur Literatursystematisierung und damit für eine literaturgestützte Anforderungsanalyse adaptiert. Dabei ist außerdem die Art der Quantifizierung, also die Berechnung und Aggregation der Kategorienhäufigkeiten, offen zu legen. Gerade bei Arbeiten, die qualitative Methoden verwenden, sollte eine möglichst sorgfältige Dokumentation der Auswertung erfolgen, damit zum einen die intersubjektive Nachvollziehbarkeit optimiert wird, und zum anderen, weil die vorliegende Arbeit in Fragestellungen eingebettet ist, die häufig mit quantitativen Methoden bearbeitet werden.

8.2.1 Anforderungsanalyse

Die Anforderungsanalyse lässt sich als eine Form der Arbeits(tätigkeits)analyse betrachten, gelegentlich werden beide Begriffe synonym verwendet. Über die Definition und Abgrenzung von Berufs-, Arbeits-, Tätigkeits-, Anforderungsanalyse, psychologischer Arbeitsstudie oder verwandten Bezeichnungen besteht in der Literatur Uneinigkeit (Frieling, 1975; Schuler, 1996). Deshalb geht diese Arbeit ähnlich wie beim Führungsbegriff von einer Arbeitsdefinition aus, die sehr allgemein gehalten wird.

Allgemein umfasst die Arbeitsanalyse eine Gruppe unterschiedlicher Verfahren zur Sammlung und Auswertung arbeitsbezogener Informationen, die sich hinsichtlich Anwendungszweck, Inhalt, theoretischer Fundierung und formaler Eigenschaften unterscheiden können. Ihre Anwendungsschwerpunkte liegen in der Eignungsdiagnostik und der Arbeitsplatzgestaltung (Schuler, 1989). In der Eignungsdiagnostik soll die Arbeitsanalyse die Verbindung zwischen Situationsdiagnostik und Personendiagnostik herstellen, indem aus ihr diagnostizierbare Personeneigenschaften abgeleitet werden. Diese hat jedoch Hypothesencharakter.

Nach Ulich (1998), der die Arbeitsanalyse insbesondere unter dem Aspekt der Gestaltung und Bewertung von Arbeitstätigkeiten betrachtet, wird ei-

ne Tätigkeitsanalyse vor allem dann durchgeführt, wenn die Ergebnisse der Auftrags- und Bedingungsanalyse nicht ausreichen, um die Arbeitstätigkeit zu bewerten und Gestaltungsvorschläge zu machen. Jede Taxonomie zur Aufgabenanalyse sollte dabei möglichst viele Informationen abbilden und eine möglichst präzise und reproduzierbare Klassifikation der Aufgaben erlauben.

Diese Klassifikationssysteme können nach Frieling (1975) inhaltlich danach unterschieden werden, auf welche Aspekte einer Aufgabe sie sich besonders oder alleine beziehen. Klassifikationen über breit gestreute Arbeitsvariablen versuchen, statt individueller Eigenschaften in erster Linie verhaltenswirksame Merkmale des Arbeitsgeschehens zu erfassen. Aufgabenklassifikationen über Arbeitsfunktionen beschreiben Aufgaben durch die nötigen Funktionen. Nach Fähigkeiten beschreibende Taxonomien stellen eine indirekte Klassifikation von Aufgaben oder Arbeitsplätzen dar, da sie auf theoretischen Schlussfolgerungen aus Eigenschaften des Arbeitsplatzes beruhen. In diesem Zusammenhang spricht man häufig von einer Anforderungsanalyse.

Auf dieser Grundlage ist für Frieling (1975) die Arbeits(platz)analyse ein Einsatz von Systematiken zur Aufgabenklassifikation sowie die praktische Nutzung der dabei erhaltenen Daten. Arbeitsanalysedaten lassen sich zu verschiedenen Zwecken nutzen. So können diese Daten als Ausgangsinformation zur Bestimmung psychischer und physischer Eignungsvoraussetzungen dienen, wie etwa zur Erstellung von Eignungsprofilen, zur Validierung eignungsdiagnostischer Testverfahren oder zur Personalauswahl und Platzierung. Weiterhin bilden Arbeitsanalysedaten die Grundlage für die Klassifikation von Berufen nach objektiven Arbeitsbedingungen oder subjektiven Anforderungen.

Berufliche Anforderungen sind eigentlich Merkmale des Arbeitsplatzes, werden aber als persönliche Leistungsvoraussetzungen behandelt, die der Stelleninhaber für die erfolgreiche Bewältigung seiner Arbeitsaufgaben mitbringen muss (Hacker, 1983). Dies umfasst Eigenschaftsanforderungen, Qualifikationsanforderungen, Verhaltensanforderungen und Ergebnisanforderungen. Eigenschaftsanforderungen sind strukturelle Merkmale einer Person, insbesondere Fähigkeiten, Persönlichkeitsmerkmale und Interessen. Unter Qualifikationsanforderungen versteht man erworbene Kenntnisse und Fertigkeiten. Verhaltensanforderungen beschreiben das notwendige Verhaltensrepertoire und kognitive Prozesse, und Ergebnisanforderungen beziehen sich auf die geforderten Lösungen und Arbeitsergebnisse.

Gerade im Führungsbereich ist es aus mehreren Gründen schwieriger als ohnehin, eine sorgfältige Anforderungsanalyse durchzuführen. Zieht man Funkes (i.V.) Systematik eignungsdiagnostischer Beschreibungsebenen heran, so wird für die Managementdiagnostik häufig die Verhaltensebene sowie die Aufgabenebene vernachlässigt. Wollte man im Führungsbereich auf der

Verhaltensebene beobachtbare Aufgabenelemente zu finden, käme man zu den Resultaten der frühen Aktivitätsstudien, die etwa die Häufigkeitsanteile von Telefonieren und Lesen angeben. Die funktionelle Bedeutung solcher Verhaltensweisen geht weit über die Aktivität per se hinaus. Zudem fehlen theoretische Modelle zur Erfassung.

Um Anforderungen einer Tätigkeit zu ermitteln, sind dennoch arbeitsanalytische Methoden hilfreich, weil sie durch die methodische und inhaltliche Vorstrukturierung die Gefahr verringern, subjektive oder falsche Vorstellungen über Führungs- beziehungsweise Managementaufgaben abzubilden. Deshalb ist es nötig, dazu ein standardisiertes Analyseverfahren einzusetzen oder selbst ein solches zu entwickeln, wozu ein theoretisches Modell benötigt wird. Eine Zielsetzung dieser Arbeit bestand darin, die Managementaufgaben mit einem Modell des komplexen Problemlösens in Verbindung zu bringen, zwischen denen es eine häufig vermutete, aber kaum systematisch untersuchte Beziehung geben soll.

Dazu sind generell zwei Zugänge denkbar, wenn man nochmals auf Funkes (i.V.) Systematik rekurriert. Auf der Aufgabenebene könnte analysiert werden, inwieweit die Situation für eine Führungskraft die gleichen Komponenten mit den gleichen Merkmalen aufweist wie ein komplexes Problem. Allerdings ist in der Problemlöseforschung die Frage, welche situativen Merkmale ein komplexes Problem charakterisieren, für diese Zwecke noch nicht befriedigend beantwortet (vergleiche Abschnitt 4.2.3).

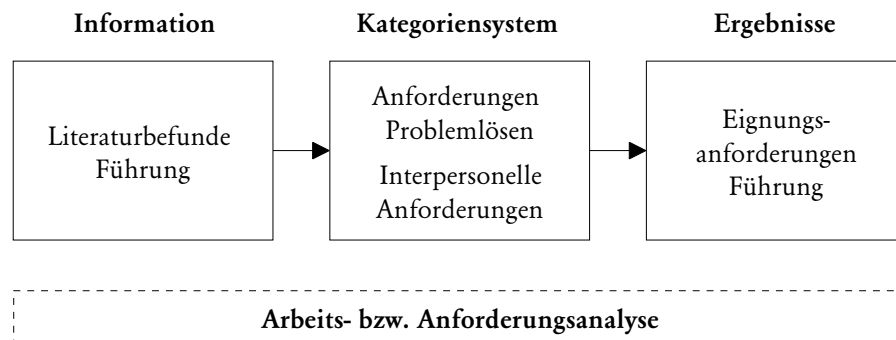


Abbildung 8.1: Ablaufschritte der Literaturanalyse als Arbeitsanalyse.

Auf der Anforderungsebene gibt es zum Problemlösen zumindest ein Modell, dass die Anforderungen, die ein komplexes Problem an den Problemlöser stellt, in einer Form abbildet, die eine Diagnose der Fähigkeit zum komplexen Problemlösen ermöglichen soll (Dörner, 1986). Ergänzt wurde dies hier um einige Anforderungen aus dem interpersonellen Bereich, da es zwar Hinweise

darauf gibt, dass Problemlösen oder andere kognitive Fähigkeiten gleichermaßen in sozialen und nicht sozialen Situationen relevant sind, dies aber noch nicht befriedigend geklärt ist. Dennoch kann dieses Vorgehen nicht nur als Anforderungs-, sondern auch als Arbeitsanalyse gesehen werden, denn es ordnet Informationen über Führungstätigkeit auf der Grundlage eines theoriebasierten Kategoriensystems.

Auf diese Weise befindet sich die Aufgabenanalyse auf der Anforderungsebene, indem Anforderungen an Führung, wie sie in der einschlägigen Literatur berichtet werden, den Anforderungen an Problemlöser (Dörner, 1986) sowie den Anforderungen im interpersonellen Bereich zugeordnet werden. Da die Modellkomponenten Bestandteile eines diagnostischen Konstrukts sind, soll dies eine Verbindung von der Analyse der Führungsaufgaben zu eignungsdiagnostisch verwertbaren Anforderungen herstellen, wobei es sich um verhaltensbezogene Anforderungen handelt.

8.2.2 Qualitative Inhaltsanalyse

Da in dieser Arbeit sprachliches Material ausgewertet wird, nämlich Arbeiten über Führung, ist eine passende Methodik für diesen Zweck zu finden. Inhaltsanalyse hat die Analyse von Material zum Ziel, das aus irgendeiner Art von Kommunikation stammt und fixiert ist, das heißt bereits protokolliert vorliegt. In Anlehnung an Mayring (1997) ist die Inhaltsanalyse eine sozialwissenschaftliche Methode mit besonderen Spezifika, die sie von anderen qualitativen Methoden unterscheidet, die sich ebenfalls mit der Analyse sprachlichen Materials beschäftigen. Diese Unterschiede liegen in mehreren charakteristischen Punkten.

So hat Inhaltsanalyse fixierte oder protokollierte Kommunikation zum Gegenstand. Sie will systematisch vorgehen, um die Einflüsse subjektiver Interpretationen auszuschließen. Diese Systematik zeigt sich darin, dass die Inhaltsanalyse regelgeleitet abläuft, um anderen Personen die Möglichkeit zu geben, dass sie die Analyse nachvollziehen und überprüfen können. Ein weiterer Aspekt des systematischen Vorgehens drückt sich darin aus, dass die Inhaltsanalyse theoriegeleitet vorgeht, also das Material vor einem Theoried Hintergrund interpretiert und damit an anderen Ergebnissen zu diesem Gegenstand anknüpft. Zuletzt will sie Rückschlüsse auf bestimmte Aspekte des bearbeiteten Materials ziehen.

Die Inhaltsanalyse hat verschiedene Aufgabenfelder. Sie dient zur Hypothesenfindung und Theorienbildung, kommt oft bei Pilotstudien vor quantitativen Verfahren zum Einsatz oder kann die Ergebnisse abgeschlossener Untersuchungen weiter vertiefen. Qualitative Analysen orientieren sich eher an Einzelfällen. Daneben werden sie für Prozessbetrachtungen eingesetzt,

um den Entwicklungsverlauf von individuellen Veränderungen festzustellen, und tragen zur Theorien- und Hypothesenprüfung bei. Die Hauptaufgabe qualitativer Forschung ist die Klassifizierung. Das bedeutet, dass Datenmaterial nach bestimmten, empirisch und theoretisch sinnvollen Gesichtspunkten geordnet wird, um das erhobene Material strukturiert zu beschreiben (Mayring, 1997). Klassifizierungen bilden entweder den Ausgangspunkt für anschließende, quantitative Analysen oder stellen selbst das Ziel der Analyse dar, zum Beispiel in Form von Typologien.

Mayring versucht, die in jeder Inhaltsanalyse enthaltenen qualitativen Bestandteile durch Analyseschritte und Analyseregeln zu systematisieren und überprüfbar zu machen. Darin lassen sich zusätzlich quantitative Schritte einbauen. Die aus den verschiedenen Quellen ausgewählten Prinzipien sollen für alle Varianten der qualitativen Inhaltsanalyse sens. Mayring (1997) gelten. Gleichzeitig will Mayring damit die Besonderheiten der Techniken aus den verschiedenen Disziplinen zu berücksichtigen. Dies erreicht Mayring dadurch, dass er drei Grundformen des Interpretierens unterteilt, die er zu insgesamt sieben Unterformen weiter aufgliedert.

Welche der sieben vorgestellten inhaltsanalytischen Formen kommen für das geplante Vorhaben in Frage? Das Textmaterial besteht in wissenschaftlichen Arbeiten, die sich aus verschiedenen Perspektiven damit auseinandersetzen, was die Anforderungen und Aktivitäten einer Führungskraft darstellen. Die Ergebnisse dieser Analysen und empirischen Studien sollen dahin gehend zusammengefasst werden, inwieweit sie sich in die verhaltensorientierte Beschreibung der vier Anforderungsdimensionen an komplexe Problemlöser nach Dörner (1986), ergänzt um interpersonelle Aspekte, einordnen lassen.

Es geht also zum einen darum, bestimmte Aspekte des Materials herauszufiltern, nämlich die Ergebnisse zur vorliegenden Frage, und zum anderen darum, das Material zu strukturieren. Dabei werden bestimmte Inhaltsbereiche unter theoretischen Gesichtspunkten extrahiert und zusammengefasst. Die theoretischen Gesichtspunkte stammen in diesem Fall aus einem Problemlösemodell und aus weiteren, diagnostisch relevanten Konstrukten. Damit ist eine inhaltliche Strukturierung angesprochen. Leider ist diese Technik bei Mayring (1997) nicht weiter ausgeführt, weshalb die Umsetzung ebenfalls hier erarbeitet werden musste.

Zudem ist es erforderlich, auf das im Text beschriebene Verständnis oder die Definition eines Begriffs zurückzugreifen, wenn geklärt werden muss, was etwa mit *Organisation* als Führungsdimension gemeint ist. Insofern ist die in erster Linie inhaltliche Strukturierung durch eine enge Kontextexplikation zu ergänzen. Das heißt, dass weitere Informationen im Text herangezogen werden, um die Bedeutung einer Kodiereinheit festzustellen oder zu rekonstruieren. Dazu müssen die Kontexteinheiten festgelegt werden, das heißt

den größten Textbestandteil, der unter eine Kategorie fallen kann und dem die Erläuterungen zugeordnet werden können.

Wird eine Kontexteinheit durch Explikationen ergänzt, gibt es zwei Varianten: Entweder wird die Bedeutung durch die Explikationen präzisiert, es erfolgt eine Monosemierung, oder die Erläuterungen dienen selbst als Indikatoren, die separat in das Kategoriensystem eingeordnet werden. In der Literaturanalyse kann beides vorkommen, denn entweder kann die Erläuterung einer Führungsdimension als Kontexteinheit direkt in eine Kategorie eingeordnet werden, oder aus der Erläuterung resultieren mehrere Kodiereinheiten, die in verschiedene Kategorien des Kategoriensystems eingeordnet werden können.

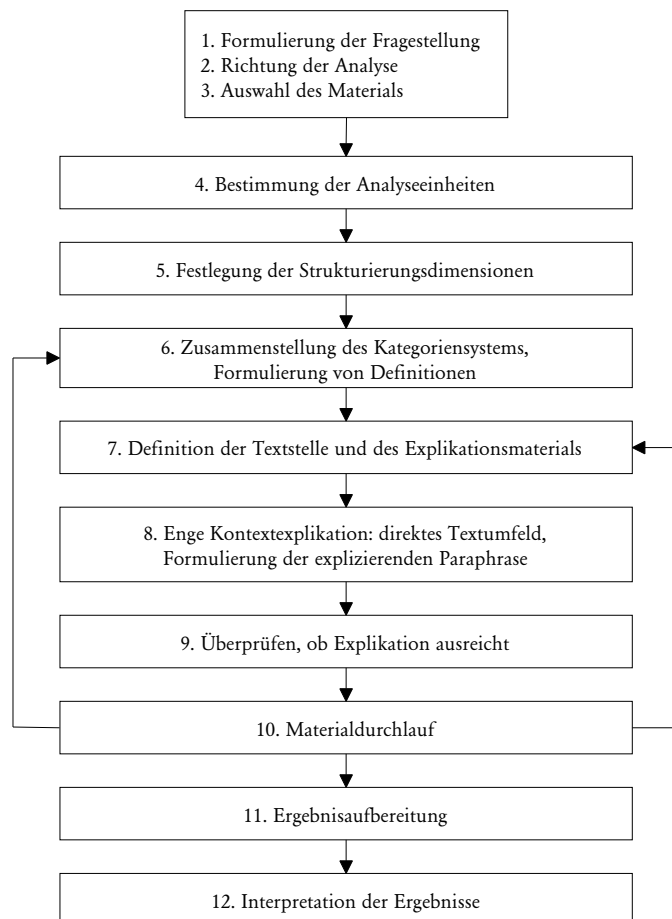


Abbildung 8.2: Eigenes Ablaufmodell der Inhaltsanalyse: inhaltliche Strukturierung mit enger Kontextanalyse.

Der eigentlichen Durchführung der Inhaltsanalyse bei einer inhaltlichen Strukturierung sind drei Schritte vorgeschaltet. So muss zunächst das Ausgangsmaterial bestimmt werden, die Fragestellung der Analyse ist zu spezifizieren, und die Analyseeinheiten müssen festgelegt werden. Sofern eine Auswahl aus einer größeren Materialmenge getroffen ist, stellt sich das Problem der Stichprobenziehung und damit der Repräsentativität. Nachdem das Ausgangsmaterial beschrieben wurde, ist eine spezifische Fragestellung zu formulieren. Warum Mayring (1997) dies erst als zweiten Schritt konzeptionalisiert, ist von einem hypothesengeleiteten Standpunkt nicht ganz nachvollziehbar, weshalb dieser Schritt des Ablaufmodells für die eigene Arbeit modifiziert wird.

Die Richtung der Analyse entscheidet, über welchen Aspekt des Textes etwas ausgesagt werden soll, denn neben dem Text selbst könnte auch die Zielgruppe, der Kommunikator oder die Quelle von Interesse sein. Hier ist die Richtung der Analyse der Text selbst. Weiterhin soll die Analyse einer theoretisch begründeten Fragestellung folgen, die sich an die bisherige Forschung über den Gegenstand anschließt. Dies wird als Kategoriensystem für die Auswertung umgesetzt, hier ist es das Anforderungsmodell an Problemlöser von Dörner (1986) ergänzt um interpersonelle Fertigkeiten.

Schließlich mussten Analyseeinheiten festgelegt werden, was vor allem für quantitative Analyseschritte und bei der Entwicklung eines Kategoriensystems wichtig ist. Dabei bestimmt die Kodiereinheit die kleinste kategorisierbare Einheit, das wäre eine von mehreren Explikationen einer Führungsdimension, die Kontexteinheit die größte kategorisierbare Einheit, eine komplette Führungsdimension. Mit der Auswertungseinheit wird festgelegt, welche Textteile ausgewertet werden, hier die Ergebnisteile und Zusammenfassungen der jeweiligen Arbeiten.

Aus den bisherigen Ausführungen ergibt sich für die eigene Fragestellung ein modifiziertes Ablaufmodell der qualitativen Inhaltsanalyse mit inhaltlicher Strukturierung nach Mayring, ergänzt um eine enge Kontextexplikation. Wie man an dem grafischen Ablaufschema aus Abbildung 8.2 sehen kann, wurden die Abfolge der Teilschritte verändert sowie einige Gesichtspunkte, welche die Charakterisierung des Materials betreffen. Das Ablaufmodell soll explizit gemacht werden, um die intersubjektive Nachvollziehbarkeit als Gütekriterium zu optimieren.

8.2.3 Kodierung

Kodiereinheiten lassen sich auf formal-syntaktischer oder auf inhaltlich-semantischer Ebene festlegen (Früh, 1998). Man kann als formale Kodiereinheiten einzelne Schlüsselbegriffe festsetzen. Immer wenn diese wörtlich im

Text erscheinen, werden sie kodiert. Daneben lassen sich Schlüsselbegriffe semantisch, also über ihre Bedeutung, definieren. Eine Aussage wird nur dann kodiert, wenn sie in der festgelegten Bedeutung auftritt. So wurde es hier praktiziert. In diesem Zusammenhang ist der Umfang der Indikatoren festzulegen, das heißt ab wann sie eine oder mehrere Kodiereinheiten darstellen. Für die Auswertung der Führungsliteratur war dies ziemlich eindeutig festzulegen, da die jeweils formulierten Bestandteile von Führung bereits untergliedert waren. So stellte die Explikation dazu einen Indikator dar beziehungsweise resultierten aus ihr mehrere Indikatoren, soweit sie sich je benanntem Aspekt in mindestens eine Kategorie einordnen ließen. Ein Explikationselement wurde höchstens einmal kodiert.

Das schließt unmittelbar an den nächsten Punkt der Kodierung an, nämlich, ob innerhalb einer Arbeit oder Auswertungseinheit dieselben Kategorien einmal oder mehrfach kodiert werden. Im Rahmen der vorliegenden Fragestellung und Auswertungsmethodik sollen gleiche Indikatoren mehrfach kodiert werden, um abzubilden, welchen Stellenwert die jeweilige Kategorie im Rahmen der Arbeit einnimmt. Außerdem ist diese Mehrfachkodierung dann darauf zurückzuführen, dass die durch das Kategoriensystem vorgegebene Abstraktionsebene etwas höher sein kann als die der Indikatoren.

Wenn ein Indikator auf der Ebene der Teilkategorien einzuordnen war, wurde er nur dort kodiert, nicht in der Hauptkategorie. Auf der Ebene der Hauptkategorie wurde er nur dann kodiert, wenn keine genauere Zuteilung möglich war. Soweit möglich, werden einzelne Analyseeinheiten auf die Ebene von Teilkategorien gebracht. Sollte dazu nicht genügend Information im Text vorliegen, werden die Analyseeinheiten explizit auf der Ebene der Hauptkategorie kodiert, was nicht ausschließt, dass weitere Analyseeinheiten auf der untergeordneten Ebene kodiert werden. So kann die anschließende Quantifizierung diese Information ebenfalls erfassen und für den Zweck einer Darstellung nur für die Hauptkategorien immer noch die Kodierung in Teilkategorien entsprechend berücksichtigen, indem diese dazu addiert oder unterteilt nach Haupt- und Teilkategorien präsentiert werden.

Nachdem ein an die Fragestellung angepasstes Ablaufmodell der qualitativen Inhaltsanalyse erarbeitet wurde, ist eine erste Form des eigenen Kategoriensystems oder einer Taxonomie zur Einordnung der Literaturbefunde im Sinne der Fragestellung zu entwickeln, was nun in Anlehnung an die beschriebenen Schritte erfolgt. Früh (1998) stellt einige Überlegungen zu diesem entscheidenden Schritt der Inhaltsanalyse an, die detaillierter als die Ausführungen von Mayring (1997) methodische Aspekte der Kategorisierung beschreiben.

Demnach soll das eigene Kategoriensystem erschöpfend oder vollständig sein, das heißt die in der Forschungsfrage vorgegebenen Materialaspekte kom-

plett erfassen. Diese sollten sich also alle in die vorher festgelegten Analyseeinheiten einordnen lassen. Gleichzeitig muss das Kategoriensystem trennscharf sein, es darf keine inhaltliche Überschneidung mehrerer Kategorien geben. Daneben müssen die Analyseeinheiten oder Indikatoren so operationalisiert sein, dass sie einerseits exklusiv oder trennscharf sind, also nicht mehrere Kategorien oder nicht interessierende Inhalte abdecken. Außerdem sollten sie die jeweils angezielte Bedeutung vollständig erfassen.

Zum möglichst eindeutigen Verständnis der Dimensionen und ihrer Teilkategorien werden diese so beschrieben, dass eine möglichst klare Zuordnung der späteren Befunde möglich ist. Weil die Befunde auf unterschiedlichen Abstraktionsebenen vorliegen, enthält auch das Kategoriensystem zwei Ebenen. Die erste Ebene, die Hauptkategorie, enthält die vier Problemlösedimensionen sowie interpersonelle Anforderungen, die zweite Ebene, die Teilkategorie, enthält die Bestandteile dieser Dimensionen. Eine minimale Analyseeinheit, also eine angegebene Aktivität, soll so spezifisch eingeordnet werden, wie dies möglich ist.

Einige Aktivitätsstudien, weniger die funktionsorientierten Arbeiten, berichten neben den allgemeinen Funktionen und Aufgaben die Fachaufgaben der jeweiligen Stichprobe. Der Zusammenhang zwischen Fachaufgaben und Führungsaufgaben wurde bereits an anderer Stelle untersucht (Ramme, 1990), aber der Vollständigkeit halber an dieser Stelle erwähnt, wobei diese Aspekte von Führungstätigkeit nicht in das Kategoriensystem aufgenommen werden. Komplexes Problemlösen, wie es bereits dargestellt wurde, bildet den theoretischen Hintergrund der folgenden vier Hauptkategorien.

Wie es in der qualitativen Inhaltsanalyse nach Mayring (1997) beschrieben ist, startet auch diese Inhaltsanalyse bei einem theoretischen Modell, in welches das analysierte Material, also die Literatur über Führungstätigkeit, eingeordnet wird. Mayring betont mehrfach, dass bei einer strukturierenden Inhaltsanalyse die grundsätzlichen Strukturierungsdimensionen genau bestimmt, aus der Fragestellung abgeleitet und theoretisch begründet werden müssen. In der Anfangsphase wird das theoretische Modell bei Bedarf in seiner kategorialen Struktur, das heißt den Dimensionen und deren Ausprägung, verändert, was wiederum auf seine Brauchbarkeit für die Klassifikation des weiteren Materials geprüft wird.

Nachdem die Kodierregeln für die Haupt- und Teilkategorien durch die Definition der inhaltlichen Dimensionen festgelegt wurden, fehlt noch ein umfassenderes Kategoriensystem, das sämtliche aufgenommenen Variablen sowie die Bezeichnung enthält. Die inhaltlichen Kategorien werden entsprechend den Ausführungen aus Kapitel 4.4 hier noch einmal in Tabelle 8.1 zusammengefasst dargestellt, die das Kategoriensystem oder die Taxonomie zur Klassifikation der Inhalte von Führungsaufgaben enthält.

Kode	Kategorie	Beschreibung
PI00	Informationsmanagement	Informationsgewinnung, -Integration
PI01	Informationsgewinnung	Strategien zur Informationsbeschaffung
PI02	Informationsintegration	Verarbeitung von Informationen
PI03	Planungsinformation	Zur Vorbereitung von Maßnahmen
PI04	Situationsinformation	Situationszusammenhänge
PI05	Kontrollinformation	Resultate eigener Maßnahmen
PI06	Systeminformation	Systemzustände
PZ00	Zielbildung	Zielausarbeitung und -balancierung
PZ01	Zielausarbeitung	Eigene Festlegung von Zielen
PZ02	Zielableitung	Ableitung von Teilzielen
PZ03	Zielgewichtung	Gewichtung von Teilzielen
PM00	Maßnahmenplanung	Beschluss von Maßnahmen
PM01	Bündelung	Bündelung von Maßnahmen
PM02	Prioritätensetzung	Festlegung der zeitlichen Reihenfolge
PM03	Effektivität	Intensität von Maßnahmen
PM04	Entscheidungsverhalten	Risikokalkulation
PM05	Anordnung	Verabschiedung der Maßnahme
PM06	Steuerung	Korrektur oder Vermeiden von Fehlern
PS00	Selbstmanagement	Emotionale Stabilität
PS01	Frustrationstoleranz	Umgang mit Misserfolgen
PS02	Belastbarkeit	Umgang mit Stressoren
PS03	Wertkonflikte	Konflikte mit eigenem Wertesystem
IP00	Interpersonelle Anforderungen	Interaktion mit anderen
IP01	Kooperation	Zusammenarbeit mit anderen
IP02	Einflussnahme	Beeinflussung anderer
IP03	Konfliktmanagement	Adäquates Verhalten in Konflikten
IP04	Kommunikation	Artikulation eigener Position
IP05	Durchsetzung	Eigene Zieldurchsetzung
IP06	Rücksicht	Berücksichtigung der Ziele anderer

Tabelle 8.1: Kategoriensystem zur Kodierung der Führungsliteratur.

Bereits öfters wurde angesprochen, dass die schließlich eingegrenzte Literatur nach ihren Befunden ausgewertet wird. Dabei ist nun im Rahmen der qualitativen Inhaltsanalyse zu konkretisieren, was genau analysiert wird. Minimale Kodiereinheit ist die Angabe einer Aktivität, Anforderung oder Funktion als Resümee oder Ergebnis der funktionalen und aktivitätsbezogenen Forschungsarbeiten. Maximale Kontexteinheit ist die jeweils vorliegende und einbezogene Forschungsarbeit. Die Auswertungseinheit oder Fundstelle ist jeweils der Ergebnis- oder Resümeeanteil der betreffenden Arbeit, der tabellarisch oder im fortlaufenden Text dargestellt wird.

Als Analyseeinheiten werden nicht die Oberbegriffe der Anforderungen an Führungs- oder Managementaufgaben herangezogen, sondern die im Text mitgelieferten Erklärungen, Definitionen oder Kurzbeschreibungen dieser An-

forderungen. Dieser Schritt zur Aufbereitung des Analysematerials ist der inhaltlichen Strukturierung im Rahmen des entwickelten Kategoriensystems vorangeschaltet, wobei es sich um eine Explikation mit einer nahen Kontextanalyse handelt. Zum einen wird dadurch die Analyseeinheit auf ein Abstraktionsniveau heruntergebrochen, das sich mit dem des Kategoriensystems erfassen lässt. Zum anderen sichert man sich dagegen ab, scheinbar gleichen Begriffen, wie beispielsweise Planung, dieselbe Definition zu unterstellen, wenn diese in verschiedenen Texten in unterschiedlicher Bedeutung verwendet werden.

8.2.4 Quantifizierung

Die Quantifizierung in der Inhaltsanalyse kommt dadurch zustande, dass für jede Forschungsarbeit die kodierten Kategorien ausgezählt, beziehungsweise ihnen Häufigkeiten zugeordnet werden. Damit findet nach der gängigen Einteilung eine Messung auf Nominalskalenniveau statt, die das geringste Messniveau darstellt, aber die Ansprüche an einen Messvorgang erfüllt. Davon abhängig ist die Festlegung zulässiger Rechenoperationen und damit verbundener statistischer Verfahren, die sich alle auf Häufigkeitsdaten beziehen.

Hier wurde eine Quantifizierungsstrategie gewählt, bei der pro Kontexteinheit oder Forschungsarbeit das Vorhandensein einer Kategorie mehrfach mit dem Wert 1 kodiert wird, unabhängig von der anteiligen Häufigkeit oder Mehrfachnennung. Damit geht vor allem bei den empirischen Untersuchungen Information verloren, und es sind außerdem die Möglichkeiten statistischer Auswertungen eingeschränkt, aber die Gefahr einer fehlerhaften Interpretation wird dadurch ebenfalls reduziert. Zudem hat diese Integration in erster Linie deskriptiven Charakter, weshalb auch der damit einhergehende Verlust an Teststärke weniger schmerzlich ist, als wenn die Ergebnisse Grundlage inferenzstatistischer Hypothesentestung bilden würden.

Weiter eingeschränkt wird die Auswahl dieser Verfahren hier durch die Stichprobengröße. So können die ausgezählten Häufigkeiten für die Gesamtheit der Arbeiten zusammengefasst und deskriptiv dargestellt werden, eine inferenzstatistische Aussage auf dieser Basis kaum möglich ist. Dies lässt sich durch eine Erweiterung der Datenbasis leicht beheben, ist also kein prinzipielles Problem der qualitativen Inhaltsanalyse, sondern eine Einschränkung dieser Arbeit. So könnten beispielsweise Zusammenhänge zwischen den Kategorienbelegungen und weiteren erfassten Eigenschaften der kodierten Arbeiten berechnet werden, wie etwa Beziehungen mit dem Erscheinungsjahr oder dem Typ der Arbeit.

Jedoch sind inferenzstatistischen Auswertungen hier ohnehin nicht angemessen, da hier vor allem ein Kategoriensystem auf Brauchbarkeit geprüft

wird. Damit handelt es sich im weiteren Sinne um die Entwicklung einer Methode beziehungsweise eines Messinstruments. Es würde den Prinzipien einer theoriegeleiteten Testkonstruktion (Conrad & Michel, 1982) widersprechen, gleichzeitig ein Instrument zu entwickeln und die Befunde auf dieser Basis im Sinne einer wissenschaftlichen Theorienprüfung inferenzstatistisch auszuwerten. Demzufolge kann über eine deskriptive Darstellung hinaus keine Schlussfolgerung gezogen werden, was die Häufigkeitsverteilung der einzelnen Kategorien bedeutet.

In dieser Phase ist lediglich eine aggregierte deskriptive Darstellung der Kategorienbelegungen sinnvoll. Die Aggregation kann dabei über die Arbeiten hinweg erfolgen, die analog zu einer konventionell-quantitativen Auswertung als Fälle behandelt werden. Weiterhin kann zusätzlich die Häufigkeit über die Teilkategorien hinweg zu den Hauptkategorien additiv zusammengefasst werden. Schließlich ist es noch möglich, die Häufigkeit zwischen den Haupt- und den Teilkategorien vergleichend darzustellen, das heißt zu betrachten, wo sich nur auf der Ebene der Hauptkategorie eine Einordnung vornehmen ließ und wo differenziertere Zuordnungen möglich waren.

8.3 Material

Anders als bei der qualitativen Inhaltsanalyse üblich, werden hier wissenschaftliche Arbeiten integriert. Deshalb fließen einige Gesichtspunkte aus der quantitativen Metaanalyse ein, die dem Problem eher gerecht werden. In der Vorbereitungsphase ist festzulegen, unter welchen Gesichtspunkten die Literatur betrachtet wird, also welche Fragestellung an sie herangetragen wird und welche Erwartungen an die Inhalte gestellt werden. Dies grenzt bereits die Art und die Menge der Literatur ein. Im nächsten Schritt ist wie bei einer Metaanalyse zu diskutieren, welche Literatur für die Auswertung herangezogen werden soll.

Wie bei der Planung des Projekts erwartet, existiert von Anfang des Jahrhunderts bis heute in der Psychologie und der Betriebswirtschaftslehre eine Fülle von Literatur, die sich aus verschiedenen Perspektiven mit Führung oder Management beschäftigt. Diese Literatur wird im Folgenden kurz charakterisiert und in ihrer Relevanz daraufhin diskutiert, wie weit sie zielführend für die formulierte Fragestellung ist. Dazu lässt sich ein metaanalytisches Qualitätskriterium für Untersuchungen, der Aspekt der Relevanz heranziehen, das ausdrückt, inwieweit die Untersuchung die Fragestellung abdeckt (Wortman, 1984).

In dieser Phase der Arbeit wird eine globale Entscheidung über Einbeziehung oder Ausschluss der jeweiligen Art von Führungsliteratur getroffen,

bevor in der nächsten Stufe an die einbezogenen Arbeiten weitere Kriterien anzulegen sind, die zu einer weiteren Selektionsentscheidung oder einer Gewichtung führen. Dabei unterscheidet sich die Literatur nach ihrem Blickwinkel auf die Thematik, sowie nach formalen, inhaltlichen und methodischen Kriterien der Wissenschaftlichkeit.

Einbezogen wurden alle Arbeiten, die prinzipiell eine begründete Antwort darauf geben können, worin die Anforderungen an Führungskräfte genau bestehen. Dabei wurden drei Kategorien gebildet. Diese sind empirische Untersuchungen über Inhalte von Führungsaufgaben, funktionale Analysen des Stellenwerts einer Führungsposition als Schnittstelle zwischen organisationalen Prozessen und Instrumente zur Beobachtung, Befragung oder Klassifikation von Führungsverhalten. Diese hier verwendeten Bezeichnungen sollen nicht die Führungsliteratur neu klassifizieren, sondern unterscheiden einbezogene Arbeiten nach hierfür relevant erachteten Gesichtspunkten.

8.3.1 Vorbereitung und Recherche

Anders als bei der qualitativen Inhaltsanalyse startet diese Arbeit nicht damit, dass bereits auszuwertendes Textmaterial vorliegt. Erst nachdem die Fragestellung und der damit verbundene theoretische Hintergrund geklärt sind, wird entschieden, welches Textmaterial in Form wissenschaftlicher Beiträge einzubeziehen ist, wobei ähnlich wie bei der Metaanalyse Selektionskriterien für die Vorauswahl bestehen und Bewertungskriterien in die Analyse mit einfließen.

Anlehnend an das bei Mayring (1997) beschriebene Kommunikationsmodell ist die Analyserichtung hier der Text selbst. Denn der Kommunikator oder die Zielgruppe als solche sind weniger interessant. Die Zielgruppe ist das Fachpublikum, der Kommunikator ein Wissenschaftler oder ein Forschungsteam. Das Thema der Zielgruppe wird im Zusammenhang mit der Auswahl des Materials kurz angeschnitten, aber letztendlich sollen nur wissenschaftliche Texte analysiert werden, die sich an ein einschlägiges Fachpublikum richten. Die Intention wird dabei im Wesentlichen die Gleiche sein, nämlich die Präsentation eigener Überlegungen und Untersuchungen aufbauend auf denen anderer Arbeiten.

Bereits eingangs wurde festgestellt, dass es der Managementdiagnostik an theoretischem Hintergrund mangelt, da nur wenig über die Anforderungen an Führungskräfte bekannt ist, was sich direkt in ein diagnostisches Konzept umsetzen lässt. Gleichzeitig bestehen sehr viele Theorien und Befunde über die Tätigkeiten oder Aufgaben von Managern. Das Ziel hier ist, diese Ansätze so zu integrieren, dass das Resultat für diagnostische Zwecke in Forschung und Praxis nutzbar ist. Als ein inhaltlich-theoretisches Rahmenkonzept zur

Integration der einschlägigen Literatur dient das Anforderungsmodell an das Lösen komplexer Probleme (Dörner, 1986), das gegebenenfalls um weitere theoretische Konzepte zu ergänzen ist, falls die Befunde der Literatur nicht vollständig eingeordnet werden können. Dabei ist zu erwarten, dass die Anforderungen an das Lösen komplexer Probleme zentral für die Aufgaben von Führungskräften sind.

Der Vorteil dieses Problemlösekonzeptes besteht in seiner Verhaltensorientierung, die gleichzeitig mit der Formulierung von Anforderungen die Grundlage für ein diagnostisches Modell bildet. Damit stellt dieses Problemlösemodell ein geeignetes Bindeglied dar zwischen der verhaltens- beziehungsweise aufgabenorientierten Führungsforschung und der auf Eignungsanforderungen angewiesenen Managementdiagnostik. Das bereits beschriebene Anforderungsmodell liegt der Entwicklung eines Kategoriensystems zur Einordnung der Literaturbefunde zugrunde, das um Anforderungen im interpersonellen Bereich erweitert wird.

Vor diesem Hintergrund lässt sich das Ziel der Arbeit auf folgende Erwartung oder Hypothese verdichten: Die Bestandteile einer Management-Tätigkeit, wie sie in der Literatur berichtet werden, lassen sich vollständig in ein Kategoriensystem einordnen, das sich in erster Linie auf das Anforderungsmodell an Problemlöser stützt. Anders ausgedrückt: Führung beziehungsweise Management enthält vergleichbare Anforderungen wie komplexe Probleme.

Ebenfalls nur in der Metaanalyse wird thematisiert, auf welchem Wege die Literatur gewonnen werden soll, und wie sicherzustellen ist, dass die schließlich zugrundegelegte Literatur eine repräsentative Stichprobe des Forschungsstandes darstellt. Sobald man sich für einschlägige Literaturdatenbanken entschieden hat, schließt das automatisch die graue Literatur aus, die eine aktuellere und möglicherweise andere Befundlage enthält. Außerdem gibt es schließlich die wegen unbefriedigender Ergebnisse unveröffentlichte Literatur, die ebenfalls einen Informationsgewinn enthält, der nie Eingang in eine Metaanalyse finden wird (*file drawer effect*). Neben den bekannten Datenbanken gibt es auch das Internet als Suchfeld, das mit kommerziellen oder fachspezifischen Suchmaschinen, Katalogen und Portalen eine weitere Möglichkeit darstellt. Hier wurde keine erschöpfende Literaturrecherche durchgeführt, wichtiger war es, eine qualitative Strategie zur Befundintegration exemplarisch zu erarbeiten, mit der hier eine spezielle inhaltliche Fragestellung verfolgt wird.

Neben inhaltlichen wurden methodische Kriterien an die Arbeiten angelegt. So muss die intersubjektive Nachvollziehbarkeit der Resultate gegeben sein, was durch sorgfältige Dokumentation des methodischen Vorgehens bezüglich Argumentation, Erhebung und Auswertung geschieht. Weiterhin

sollte die Methodik für den jeweiligen Forschungsstand bezüglich der abgedeckten Inhaltsbereiche, Literatur, Erhebungs- und Auswertungsverfahren angemessen sein. Die Stichprobengröße wird als wichtiges Qualitätskriterium betrachtet, auch wenn damit etliche Untersuchungen wegfallen, da zum Beispiel Einzelfallstudien sehr detailliert sein können, aber keine Information über die Repräsentativität der Befunde liefern. Schließlich schränken Störfaktoren beziehungsweise methodische Fehler den Geltungsbereich der Arbeit im Sinne der Fragestellung ein.

Arbeiten, welche diesen methodischen Kriterien nicht genügen, sollen von der Auswertung völlig ausgeschlossen werden. Bei der Metaanalyse hat man die Wahl, ob man die methodischen Eigenschaften und damit Qualitätsunterschiede kodiert und als Moderator in die Auswertung mit aufnimmt, um damit Varianz der Ergebniswerte aufzuklären, oder ob die methodisch problematischen Studien weggelassen werden. Da es bei dem anvisierten Material zunächst kaum möglich ist, Varianz zu berechnen, geschweige denn aufzuklären, wäre dies hier ohnehin nicht möglich. Ähnliches gilt für die Variante, die Studien danach zu gewichten.

8.3.2 Ausgeschlossene Literatur

Zunächst wird kurz neben den bereits diskutierten Führungsansätzen Literatur beschrieben, die aus der Analyse ausgeschlossen wurde. Da sich diese Arbeit aus wissenschaftlicher Perspektive mit dem Inhalt und den Anforderungen von Führungstätigkeit auseinander setzt, sollten *populärwissenschaftliche Literatur und Ratgeber* eigentlich gar nicht behandelt werden. Allerdings repräsentiert diese Literatur stärker die Perspektive der Stakeholder, der praktisch Tätigen in der Managementdiagnostik. Letztere sind bei ihrer Arbeit häufig mit Erwartungen konfrontiert, deren Basis die so genannte populärwissenschaftliche Literatur darstellt. Deshalb wird dieses Genre von Führungsliteratur hier kurz vorgestellt, und begründet, warum es aus der vorliegenden Arbeit ausgeschlossen wird.

Populärwissenschaftliche Literatur will wissenschaftliche Ansätze und Befunde auf verständliche Art und Weise an ein Laienpublikum vermitteln, ohne dass dazu einschlägiges Vorwissen oder die Beherrschung fachspezifischer Termini notwendig ist. Dazu entfernt sie sich in unterschiedlichem Ausmaß von der methodischen und inhaltlichen Basis der wissenschaftlichen Arbeiten. Weiterhin genügt sie eher selten den formalen Kriterien wissenschaftlicher Arbeiten, wie etwa dem Verweis auf die Quellen und die empirischen Grundlagen der präsentierten Behauptungen und Resultate.

Sofern eine solche Arbeit lediglich eine Art Übersetzung der wissenschaftlichen Literatur für Laienpublikum darstellt, ist gegen seine Verwertung in

dieser Arbeit prinzipiell nichts einzuwenden, doch kann man dann genauso die wissenschaftliche Originalliteratur verwenden. Beruht sie jedoch auf einer Zusammenfassung verschiedener Ansätze oder auf persönlichen Erfahrungen, deren einzelner Stellenwert für die Argumentation meist nicht mehr zu trennen ist, schränkt dies die Brauchbarkeit für den vorliegenden Zweck extrem ein. Genau genommen ist sie unter diesen Umständen auch für das Laienpublikum nicht brauchbar.

Von populärwissenschaftlicher Literatur zur *Ratgeberliteratur* gibt es einen fließenden Übergang. Ratgeberliteratur hat weniger den Anspruch, den Stand der Forschung wiederzugeben, sondern gibt vielmehr generelle oder spezifische Handlungsempfehlungen für optimalen Führungserfolg. Solche Empfehlungen werden anekdotisch oder gar nicht begründet und selten mit empirischen Studien fundiert. In der Regel betrachtet sie auch das Erfolgskriterium nicht systematisch. Damit besteht die Relevanz populärwissenschaftlicher Veröffentlichungen und von Ratgeberliteratur lediglich in ihrer großen Verbreitung. Offen bleibt, ob sie mit ihren Inhalten Common-sense-Meinungen reflektiert oder für diese verantwortlich ist. Deshalb wurde sie hier zumindest thematisiert und gleichzeitig von einer weiteren Verwendung explizit ausgeschlossen.

Präskriptive Managementliteratur behandelt Führungsmodelle, die in Form von Handlungsanweisungen Orientierung vermitteln und Empfehlungen geben wollen, wie Führungskräfte die Bedürfnisse ihrer Mitarbeiter berücksichtigen und beeinflussen können, um auf deren Leistungsverhalten einzuwirken. Diese Modelle enthalten strukturelle und personale Elemente. Strukturelle Elemente machen Vorgaben über die organisatorische Struktur und Regelungen für optimale Führung. Die personalen Elemente beziehen sich auf Aspekte der Interaktion, wie Führungsverhalten, Führungsstil oder Beziehung zwischen Vorgesetzten und Mitarbeitern.

Damit lassen sich weiterhin Partial- von Totalmodellen unterscheiden, da Partialmodelle nur einzelne strukturelle oder personale Bausteine enthalten, während Totalmodelle beides abdecken wollen. Unter den Totalmodellen wurde vor allem das Harzburger Modell (Höhn, 1974) oder der Ansatz des Management by Objectives (Drucker, 1954) bekannt. Partialmodelle setzen an einzelnen wahrgenommenen Defiziten des Führungsalltags an, aus denen rezeptartige Empfehlungen abgeleitet werden, beispielsweise Management by Motivation.

Gemeinsam ist diesen unterschiedlich weit reichenden Ansätzen, dass sie häufig praktiziert, aber selten systematisch auf ihre Wirksamkeit untersucht werden. Ebenso fehlt ihnen zumeist eine wissenschaftliche theoretische oder argumentative Basis für die postulierten Annahmen und daraus abgeleiteten Empfehlungen. Selbst wenn dies gegeben wäre, ist aber daraus noch nicht

abzuleiten, welche Anforderungen der Arbeitsalltag an eine Führungskraft stellt, sondern lediglich, welcher Techniken sie sich bedienen sollte. Aus diesen Gründen wird die präskriptive Managementliteratur, die sich zudem mit der beschriebenen Ratgeberliteratur stark überschneidet, ebenfalls nicht weiter für diese Arbeit berücksichtigt.

8.3.3 Funktionale Analysen

Die funktionale Managementliteratur bildet die Arbeit von Führungskräften als Ausübung von Funktionen nach, die unabhängig von den einzelnen Führungspersonen sind. Dabei wird Management als besondere Funktion angesehen, die in arbeitsteiligen Organisationen nötig ist. Sie führt Steuerungsprozesse innerhalb und zwischen Organisationen aus, um das Weiterbestehen der Organisation als System zu gewährleisten. Damit erfüllen Manager ähnliche Funktionen wie andere organisationale Prozesse. In diesem Rahmen, der implizit oder explizit Organisationen aus einer systemtheoretischen Sichtweise begreift, sind Organisationen übergeordnete Funktionseinheiten, die sich aus kleineren Organisationseinheiten zusammensetzen, von denen eine die Administration ist, welche die Managementfunktionen beinhaltet.

Dabei werden Daten meist durch schriftlicher Befragungen mit strukturierten und weitgehend standardisierten Fragebögen erhoben. Sie lassen die Häufigkeit und Wichtigkeit von Verhaltensweisen einschätzen, die den jeweiligen Funktionsbereichen zuzuordnen sind. Die Auswertung stützt sich meist auf Häufigkeitsverteilungen, um den Anteil der Arbeitszeit für die einzelnen Funktionen zu ermitteln. Manchmal werden auch strukturentdeckende Verfahren, in der Regel Faktorenanalysen, durchgeführt, um die Klassifikationen zu modifizieren. Die Mehrzahl der Studien arbeitet mit vorgegebenen Kategorisierungen aus der Literatur.

Diese Art von Literatur soll aus mehreren Gründen einbezogen werden: Zum einen ist die funktionale Betrachtungsweise von einzelnen Personen unabhängig und ergänzt dadurch die sozialwissenschaftlich-psychologischen Befragungen und Verhaltensbeobachtungen aus den Aktivitätsstudien und Kategoriensystemen, die einen realistischen, aber möglicherweise defizitären Status quo abbilden, um einige Aspekte eines theoretisch sinnvollen Führungshandelns. Zum anderen wird sie durch empirische Befunde unterstützt, die in Details voneinander abweichen, aber insgesamt ein konsistenteres Bild liefern als etwa die Ergebnisse von Führungstheorien.

Solche Beiträge stammen meist aus der betriebswirtschaftlich ausgerichteten Management-Literatur, weniger aus der sozialwissenschaftlichen Führungsliteratur. Funktionale Arbeiten können Artikel in Fachzeitschriften sein, Monografien oder Management-Lehrbücher. Letztere formulieren ihr

Verständnis von Management nicht explizit als Modell, sondern legen ein solches implizit der Gliederung des Buches zugrunde.

-
- Barnard (1968)** The Function of the Executive. Cambridge, Mass: Harvard University Press
Baugut und Krüger (1976) Unternehmensführung: Modelle - Strategien - Techniken. München: Opladen
Fayol (1929) Allgemeine und industrielle Verwaltung. Berlin
Goodman (1968) A System Diagram of the Functions of a Manager. California Management Review
Koontz und Wehrich (1993) Management: A Global Perspektive. McGraw-Hill
Miner (1973) The Management Process – Theory, Research and Practice. New York
Newman (1963) Administrative Action. New Jersey: Englewood Cliffs
Parsons (1960) Structure and process in modern society. New York
Terry und Franklin (1982) Principles of Management. Homewood
Urwick (1956) The Patterns of Management. Minneapolis: University of Minnesota Press
Wren und Voich (1984) Management: Process, Structure and Behavior. New York: Wiley
Yukl (1994) Leadership in Organizations. New Jersey: Prentice Hall
-

Tabelle 8.2: Übersicht über die verwendeten funktionalen Analysen.

Diese Art von Literatur enthält keinen Ergebnisteil, da ihre Darstellung ja nicht dem Ablaufschema einer empirischen Untersuchung folgen kann, sondern meist zusammenfassende Überblicksteile am Anfang oder Ende der Arbeit. Diese überblicksartigen Darstellungen wurden im Rahmen der Inhaltsanalyse ausgewertet und soweit durch Explikationen ergänzt, bis eine Einordnung in das Kategoriensystem möglich war. Auch hierbei bewährte sich das Kategoriensystem gut, da alle Dimensionen, Anforderungen oder Prinzipien von Management mit den erarbeiteten Kategorien in Deckung zu bringen waren.

Zur Erfassung der funktionalen Arbeiten wird neben den üblichen Angaben der Hintergrund dokumentiert, das heißt ob es sich um einen Beitrag in einer Fachzeitschrift, eine Monografie oder ein Lehrbuch handelt, außerdem, worin der Anspruch der Arbeit besteht und worauf sich dieser stützt. Ein Lehrbuch erhebt einerseits den Anspruch, Grundlagenwissen zu vermitteln und bisherige Ansätze strukturiert darzustellen, eventuell auch zu bewerten, behauptet aber nicht explizit, damit ein eigenes und neuartiges Verständnis von Führung oder Management zu präsentieren.

Bereits bei der exemplarischen Inhaltsanalyse zeigte sich, dass sich die Auffassung auch bekannter Einteilungen von Führungsfunktionen im Einzelnen unterscheiden. Deshalb werden funktionale Arbeiten mit Lehrbuchcharakter einbezogen, auch wenn Carroll und Gillen (1987) nach Durchsicht von 21 Managementlehrbüchern feststellten, dass 17 davon ihr Buch nach den Fayol'schen Managementfunktionen gliederten. Als Monografie bezeichnete Arbeiten, im engeren Sinne als wissenschaftliche Behandlung eines einzelnen Themas definiert, sollen hier als Buch veröffentlichte Arbeiten charakterisieren, deren Anspruch über die Zusammenfassung bekannter funktionaler Ansätze und Befunde zur Führung hinausgeht, weil sie explizit eine eigene Einteilung, Konzeptionalisierung oder Sichtweise von Führungs- beziehungsweise Managementfunktionen enthalten.

8.3.4 Empirische Untersuchungen

Eine völlig andere Perspektive tragen die empirischen Untersuchungen zu der beschriebenen Problemstellung bei. Sie beschreiben möglichst konkret, was Manager täglich tun. Dies stellte eine Reaktion auf die bis dahin durchgeführten analytisch-funktionalen Beschreibungen des Managementprozesses dar. Um die Datenbasis zu erweitern, soll eine möglichst genaue und umfassende fortlaufende Dokumentation des Verhaltensstroms im alltäglichen Arbeitsvollzug erfolgen, um damit eine breitere und fruchtbare empirische Basis zur Theorienbildung zu erhalten. Von Aktivitätsbeschreibungen werden wichtige Anregungen für die Lösung praktischer Probleme wie Training und Auswahl (Stewart, 1976) erwartet, wofür funktionale Analysen ungeeignet seien.

Dazu nutzt man hauptsächlich Verfahren der Selbst- oder Fremdbeobachtung sowie mündliche oder schriftliche Befragung. Diese Methoden kommen in älteren Arbeiten einzeln, später auch kombiniert zum Einsatz, um eine facettenreichere Beschreibung des Arbeitsverhaltens zu ermöglichen und gleichzeitig die speziellen Schwächen der einzelnen Methoden zu kompensieren. Die Auswertung erfolgt in der Regel mit qualitativen und quantitativ-deskriptiven Verfahren, nur selten mit multivariaten statistischen Analysen. Zum einen geht es darum, die Häufigkeit, Dauer und äußeren Umstände der Aktivitäten festzuhalten, zum anderen interessiert, zu welchem Zweck Führungskräfte arbeiten.

Für die Auswertung im Rahmen dieser Arbeit sollen empirische Untersuchungen ein Kernstück bilden, weil sie eine umfassende empirische Basis zu der Frage anbieten, worin die Tätigkeiten von Führungskräften bestehen. Das bildet den Ausgangspunkt für weitere Überlegungen, aus denen theoretisch schließlich ein Anforderungsmodell mit Zielrichtung auf diagnosti-

zierbare Eigenschaften resultieren kann. Allerdings sollen nicht alle Aktivitätsstudien einbezogen werden, sondern nur solche, die bereits Informationen darüber mitliefern, worin der Zweck oder die Funktion der einzelnen Tätigkeiten besteht. Die Informationen einer Studie, die lediglich die Dauer, Häufigkeit oder den zeitlichen Anteil von Telefonieren, Schreibtischarbeit oder Ähnlichem berichtet, können nicht in Bezug zu einem verhaltensorientierten Anforderungsmodell gesetzt werden. Dieser Aspekt der Abstraktionsebene grenzte die Literaturstichprobe weiter ein.

Allan (1981) Managers at work: A large scale study of the managerial job in New York City Government. Academy of Management Journal (1476 Führungskräfte)

Haas, Porat und Vaughan (1969) Actual vs. ideal time allocations reported by managers: A study of managerial behavior. Personnel Psychology (355 Führungskräfte)

Hinrichs (1975) Where Has All the Time Gone? Personnel (142 Führungskräfte)

Horne und Lupton (1975) The work activities of middle managers - an exploratory study. The Journal of Management Studies (66 Führungskräfte)

Kotter (1982) What effective managers really do. Harvard Business Review (15 Führungskräfte)

Lau, Newman und Broedling (1980) The nature of managerial work in the public sector. Public Management Forum (370 Führungskräfte)

Mahoney, Jerdee und Carroll (1965) The Job(s) of Management. Industrial Relations (425 Führungskräfte)

Mintzberg (1973) A New Look at the Chief Executive's Job. Organizational Dynamics (5 Führungskräfte)

Whitely (1985) Managerial work behavior: An integration of results from two major approaches. Academy of Management Journal (70 Führungskräfte)

Tabelle 8.3: Übersicht über die verwendeten empirischen Studien (N Gesamt = 2824).

Kennzeichnend für hier so bezeichnete empirische Studien ist die Erhebung empirischer Primärdaten bei Führungskräften über deren Aktivitäten oder Anforderungen im Arbeitsalltag, wobei sie sich bei der Erhebung entweder auf vorhandene Konzepte stützen oder die Befunde induktiv in diesem Sinne verdichten. Wichtig war bei der Auswahl weiterhin, dass die Befunde eine Aussage auf einem vergleichbaren Abstraktionsniveau treffen wie das theoretische Rahmenmodell, in das sie eingeordnet werden. Ein zu hohes Abstraktionsniveau der Resultate stellte weniger ein Problem dar, da sich diese anhand der zusätzlichen Erläuterungen, Definitionen oder Beschrei-

bungen aus der Arbeit in der Regel in das Modell einordnen ließen. Zur Kodierung der empirischen Studien über Anforderungen, Aufgaben oder Aktivitäten von Führungskräften werden mehrere Informationen herangezogen. Neben einer kurzen Darstellung der Fragestellung werden für jede Studie die Art und der Umfang der Stichprobe sowie die verwendeten Erhebungs- und Auswertungsverfahren erfasst.

8.3.5 Instrumente

Zwischen empirischen Untersuchungen und funktionalen Analysen nehmen die Ansätze, die hier als Erhebungsinstrumente bezeichnet werden, in mehrfacher Hinsicht eine Mittelstellung ein. Zum einen werden sie in Studien beider Ansätze zur Erhebung eingesetzt, zum anderen haben sie oft ein höheres Aggregationsniveau als Aktivitätsstudien und gleichzeitig sind ihre Inhalte spezieller und konkreter als die von funktionalen Betrachtungen.

Wie so oft, ist diese Abgrenzung von Instrumenten gegenüber den anderen dargestellten Arbeiten künstlich und willkürlich, und es bestehen Zusammenhänge sowie fließende Übergänge. So basieren die Dimensionen und Definitionen von Instrumenten, häufig auch als Kategoriensysteme bezeichnet, teilweise auf funktionalen Analysen, teilweise auf induktiven Verdichtungen von Verhaltensbeobachtungen. Da aber zu ihrer Entwicklung eigene empirische Studien durchgeführt werden, sollen Kategoriensysteme hier unter einer getrennten Rubrik aufgeführt und als solche für die spätere Auswertung berücksichtigt werden.

Kategoriensysteme ermöglichen die Beobachtung von Verhalten, indem sie mögliche Verhaltensweisen benennen und klassifizieren. Sie werden in der Regel im Zusammenhang mit einem bestimmten theoretischen und methodischen Ansatz der Führungsforschung entwickelt, entweder um Beobachtungen in diesem Rahmen zu strukturieren, oder sie sollen auf dem theoretischen Hintergrund Diagnosen des Führungsverhaltens ermöglichen. Damit müssen solche Kategoriensysteme den bekannten diagnostischen Gütekriterien genügen, auf die das jeweilige Verfahren explizit geprüft werden kann, wie auf Reliabilität und Validität. Die Validität wird in der Regel als Kriteriumsvalidität bestimmt, also über den Zusammenhang mit Führungserfolg.

Üblicherweise hat ein solches System einen dimensional Aufbau, das heißt, entsprechend der theoretischen Basis sind die Items oder Beobachtungseinheiten Teilbereichen des Verhaltens zuzuordnen, die überschneidungsfrei und vollständig sein sollten, um abhängige Daten zu vermeiden. Als Erhebungsinstrumente werden alle Fragebogenverfahren, Kategoriensysteme und Taxonomien bezeichnet, die dazu dienen, Führungsverhalten zu erfassen oder sonstige empirische Informationen über Führung einzuteilen. Sie dienen

als Verfahren zur Tätigkeitsanalyse dazu, die Inhalte von Führungstätigkeit zu beschreiben und gleichzeitig die verhaltensbezogenen Anforderungen für eine erfolgreiche Tätigkeitsausübung zu identifizieren. Für die inhaltsanalytische Auswertung wurden die Skalen, Kategorien oder Dimensionen des jeweiligen Verfahrens herangezogen.

-
- Brooks (1950)** What Successful Executives Do. Personnel
Luthans und Lockwood (1984) Towards an observation system for measuring leader behavior in natural settings. In Hunt, Hosking, Schriesheim & Stewart (Hrsg.), Leaders and managers: International perspectives on managerial behavior and leadership. New York: Pergamon Press.
Morse und Wagner (1978) Measuring the Process of Managerial Effectiveness. Academy of Management Journal
Page (1978) Managerial Position Description Questionnaire
Stogdill (1963) Manual for the Leader Behavior Description Questionnaire - Form XII. Columbus: Ohio State University
Yukl (1989) Managerial leadership: A review of theory and research. Journal of Management
-

Tabelle 8.4: Übersicht über die verwendeten Erhebungsinstrumente.

8.4 Ergebnisse

Bis zu diesem Punkt wurde eine Strategie erarbeitet, wie sich mit Techniken der qualitativen Inhaltsanalyse die eingegrenzte Führungsliteratur im Sinne einer Anforderungsanalyse auswerten lässt. Ansätze zum Problemlösen sowie zu interpersonellen Anforderungen liefern die Grundlage für ein theoriebasiertes Kategoriensystem, das eine inhaltliche Strukturierung der Literaturbefunde zu der Frage ermöglicht, welche Inhalte Führungsaufgaben haben. Außerdem ist eine Entscheidung gefallen, welche der in Frage kommenden Literatur einbezogen wurde und welche nicht. Im Probelauf wird der Vorgang der Kodierung ausführlich und beispielhaft dargestellt, damit dies prinzipiell für die gesamte Literaturstichprobe nachvollziehbar wird.

Nach dem Probedurchlauf kommt das Ablaufmodell für die literaturgestützte Anforderungsanalyse im Hauptlauf zum Einsatz. Die Ergebnisdarstellung gliedert sich auf in die Übersicht über die zusammengefassten Häufigkeiten für alle Arbeiten zusammen sowie in die Aufgliederung nach den drei unterschiedenen Arten einbezogener Literatur. Weiterhin werden diese Ergebnisse für die fünf Hauptkategorien global sowie für die Haupt- und Teilka-

tegorien der fünf Bereiche des Kategoriensystems unterteilt präsentiert. Eine inferenzstatistische Testung dieser Unterschiede ist bei der geringen Stichprobengröße jedoch ausgeschlossen, sondern das vorliegende Material kann nur deskriptiv beschrieben werden.

Diese Verteilungen lassen sich unter zwei Gesichtspunkten betrachten. Zum einen stellt sich die Frage, wie gut das Kategoriensystem ausgenutzt wird, zum anderen, ob sich alle Analyseeinheiten einordnen und damit kodieren lassen. Wenn das Kategoriensystem ausgenutzt wird und wenig zweifelhafte Kodiereinheiten übrig bleiben, bedeutet das noch nicht, dass dies die einzige Möglichkeit ist, die Führungsliteratur in allen Bedeutungsaspekten erschöpfend zu systematisieren, aber zumindest gibt dies einen Hinweis darauf, dass eine solche Strukturierung eine sinnvolle Möglichkeit darstellt.

8.4.1 Probelauf

Mayring (1997) empfiehlt, in einem Probelauf zunächst eine Teilstichprobe des gesamten Materials zu kategorisieren, um daran die Brauchbarkeit der Haupt- und Teilkategorien zu erproben sowie das Explikationsprinzip der Analyseeinheiten zu testen und beides gegebenenfalls zu korrigieren. Mit dem endgültigen Kategoriensystem werden die restlichen Forschungsarbeiten kodiert, anschließend quantifiziert und diese quantitativen Daten deskriptiv ausgewertet. Der erste Materialdurchlauf einer strukturierenden Inhaltsanalyse dient dazu, das Kategoriensystem so sehr wie nötig zu erweitern oder zu präzisieren.

Das Kategoriensystem, die erarbeitete Methodik und das Kodierungsschema können unter verschiedenen Aspekten erprobt werden. Dabei wird auch der Kodierleitfaden erweitert, indem besonders eindeutige Zuordnungen als Ankerbeispiele aufgenommen und bei uneindeutigen Einschätzungen zusätzliche Kodierregeln formuliert werden. Hier werden zunächst exemplarisch drei Arbeiten aus den drei ausgewählten Bereichen der Führungsforschung kodiert, um zum einen die Vorgehensweise zu veranschaulichen und zum anderen Ankerbeispiele für die weitere Kodierung zu erhalten. Deshalb dient die Darstellung des Probelaufs außerdem dazu, die Kodierung für drei exemplarische Arbeiten etwas genauer auszuführen. Jede Arbeit repräsentiert eine der drei Literaturkategorien.

Für eine Arbeit wurde zunächst die Hauptkategorie oder Dimension. Deren Beschreibungen wurden im Rahmen der Kontextexplikation mit aufgenommen und solange in mehrere Aspekte aufgeteilt, bis eine minimale Kodiereinheit gewonnen wurde, die sich in genau eine Teilkategorie eindeutig einordnen ließ. Falls die Explikation nicht mehr hergab, wurde die Kodiereinheit auf der Ebene der Hauptkategorien eingeordnet.

Diese Kodiereinheiten wurden den Haupt- oder Teilkategorien des Kategoriensystems zugeteilt und schließlich wurde das Variablenlabel für die Übernahme in die Statistik-Software angegeben. Die Beschreibungen der zumeist englischsprachigen Arbeiten werden hier gleich in der deutschen Übersetzung wiedergegeben, um die Zuordnung zur jeweiligen Kategorie für den Leser und den Kodierer besser nachvollziehbar zu machen. Das Ergebnis dieses Vorgangs veranschaulichen die im Folgenden dargestellten Tabellen für jede Arbeit. Die Analyseeinheit ist die Bezeichnung der Führungsdimension, die Kodiereinheit ein kodierbares Element der Explikation und die Kategorie die Bezeichnung der jeweiligen Haupt- oder Teilkategorie, die schließlich kodiert wurde. Die Abkürzung für den Code entspricht der Darstellung in Tabelle 8.1.

Analyseeinheit	Kodiereinheit	Kategorie	Code
Subgoal Setting	Teilziele vorgeben	Zielableitung	PZ02
Organizing	Aufgaben an Personal zuweisen	Maßnahmenbündelung	PM01
Allocating	Aufgaben zu Zielen zuordnen	Maßnahmenplanung	PM00
Systematizing	Beziehung zwischen Systemteilen feststellen	Situationsorientierung	PI04
Plan	Planung	Maßnahmenplanung	PM00
Leading	Mitarbeiterführung	Beeinflussen	IP02
Motivation	Erreichen, dass Mitarbeiter arbeitet	Durchsetzung	IP03
Directing	Veranlassen der Arbeitsabläufe	Anordnung	PM05
Training	Unterweisen	Kommunikation	IP04
Monitoring Utilization	Überwachen von Ressourceneinsatz	Effektivität	PM03
Comparing	Abgleich Input–Output	Kontrollinformation	PI05
Monitoring Progress	Fortschritte überwachen	Systeminformation	PI06

Tabelle 8.5: Exemplarische Kodierung der ersten funktionalen Arbeit (Goodman, 1968).

Die erste funktionale Arbeit, die kodiert wurde, stammt von Goodman (1968). Sie stellt eine Erweiterung des Ansatzes von Fayol dar. Dieser Beitrag will ein integratives Rahmenkonzept funktionaler Managementansätze vorstellen, das außerdem systemtheoretische Überlegungen enthält, indem die Komponenten Input, Planung, Ausführung und Kontrolle weiter aufge-

gliedert und in einem geschlossenen Regelkreis aufeinander bezogen werden. In die inhaltsanalytische Auswertung flossen jedoch nur die einzelnen Funktionen und Teilfunktionen sowie deren Explikationen ein, die Beziehungen im Rahmen des Regelkreises blieben ausgeschlossen.

Analyseeinheit	Kodiereinheit	Kategorie	Kode
Supervision of Employees	Arbeit anweisen	Beeinflussen	IP02
	Unterweisen	Kommunikation	IP04
	Sanktionen	Durchsetzung	IP05
Harmonizing	Bei Spannungen vermitteln	Konfliktmanagement	IP03
Information handling	Information beschaffen	Informationsgewinnung	PI01
	Information verarbeiten	Informationsintegration	PI02
	Information weitergeben	Kommunikation	IP04
Analytical-evaluative	Gesetze, Berichte, Dokumentationen analysieren	Situationsinformation	PI04
Change-Initiating	Systemveränderungen veranlassen	Anordnung	PM05
Monitoring	Fortschritte feststellen	Kontrollinformation	PI05
	Aktivitäten beobachten	Systeminformation	PI06

Tabelle 8.6: Exemplarische Kodierung der ersten empirischen Studie (Allan, 1980).

Von Allan (1980) stammt die erste kodierte empirische Studie, in der bei Führungskräften aller Managementebenen und Funktionen mit einem Fragebogen die Aktivitäten, deren Häufigkeit und Bedeutung erhoben und faktorenanalytisch reduziert wurden. Diese wurden mit einer Bezeichnung und charakterisierenden Beschreibung dargestellt. Ziel der Studie war es, Aufgaben von Führungskräften zu diagnostizieren und induktiv zusammenzufassen. Dazu wurde in einer kurz dokumentierten Vorstudie ein Fragebogen mit 146 Items in der Endfassung entwickelt, wobei die Items jeweils eine Teilaufgabe enthalten, deren Dauer und Bedeutung auf einer Ratingskala von null bis vier einzuschätzen waren. Dieser wurde an insgesamt 1476 Führungskräften aller sechs Ebenen der Stadtverwaltung von New York eingesetzt. Nur die 57 Aufgaben, deren durchschnittliche Bedeutsamkeit und Dauer jeweils einen Cutoff-Wert erreichte, wurden einer Faktorenanalyse unterzogen, aus der sechs Dimensionen resultierten, wie sie mit ihrer Beschreibung in der Tabelle dargestellt wurden.

Das bekannte Leader Observation System (LOS) von Luthans und Lockwood (1982) stellt eine Systematik zur Verhaltensbeobachtung zur Verfü-

Analyseeinheit	Kodiereinheit	Kategorie	Kode
Planning, Coordinating	Ziele festlegen	Zielbildung	PZ00
	Aufgaben verteilen	Maßnahmen- bündelung	PM01
	Reibungslosen Ablauf gewährleisten	Prioritäten- setzung	PM02
Staffing	Arbeitsbeschreibungen formulieren	Kommunikation	IP04
	Bewerber auswählen	Entscheidungs- verhalten	PM04
Training, Development	Mitarbeiter ausbilden	Kommunikation	IP04
Decision Making, Problem solving	Probleme definieren	Planungs- information	PI03
	Zwischen alternativen Strategien entscheiden	Entscheidungs- verhalten	PM04
	Prozeduren zur Effizienz- steigerung entwickeln	Effektivität	PM03
Processing Paper- work	Post, Berichte lesen	Informations- management	PI00
	Berichte, Memos, Briefe schreiben	Kommunikation	IP04
Exchanging Routine Information	Information erhalten	Informations- management	PI00
	Information weitergeben	Kommunikation	IP04
Monitoring, Controlling Performance	Arbeitsabläufe überwachen	Steuerung	PM06
	Leistungsdaten beobachten	Kontroll- information	PI05
	Instandhaltung	Steuerung	PM06
Motivation, Reinforcing	Belohnung, Feedback	Einflussnahme	IP02
	Delegation	Maßnahmen- anordnung	PM05
Disciplining, Punishing	Diverse Sanktionen	Durchsetzung	IP05
Interaction with Outsider	Kunden, Öffentlichkeit	Soziale Interak- tion	IP00
Managing Conflict	In Konflikte eingreifen	Konflikt- management	IP03
Socializing, Politicking	Tratsch, Intrigen	Kommunikation	IP04

Tabelle 8.7: Exemplarische Kodierung des ersten Erhebungsinstruments (Luthans & Lockwood, 1982).

gung. Es wurde exemplarisch für Erhebungsinstrumente kodiert. Bei der Entwicklung des Leadership Observation System (LOS) sollte eine Systematik zur Verhaltensbeobachtung von Führungstätigkeiten geschaffen werden, die analoge Inhalte hat wie vergleichbare Fragebogenverfahren.

Dazu wurden zunächst 44 Führungskräfte über zwei Wochen insgesamt zehn Stunden lang beobachtet und diese Beobachtungen in einem unstrukturierten Protokoll festgehalten. Aus der wiederholten Anwendung der Delphi-Technik mit Experten und Laien resultierten schließlich die zwölf Kategorien des LOS. Luthans und Lockwood bestimmten in einer weiteren Studie an 120 Führungskräften die Reliabilität und die Validität des LOS. Letzteres erfolgte im Rahmen einer Multitrait-Multimethod-Validierung mit zwei Führungsfragebögen.

Bei der Kodierung konnten sämtliche Kodiereinheiten in der Systematik untergebracht werden konnten, und es gab auch keine Unklarheiten über die Zuordnung. Dies wurde erleichtert durch den zusätzlichen Schritt der engen Kontextexplikation. Immer dann, wenn in der Spalte neben einer Analyseinheit mehrere Kodiereinheiten aufgeführt und kategorisiert sind, liegt dieser zusätzliche Schritt vor. Es zeigte sich, dass sich die Definitionen von gleichen oder ähnlichen Bezeichnungen für Führungs- oder Managementaufgaben in Abstraktionsgrad und Inhalt unterscheiden. So beinhaltet beispielsweise Monitoring bei Allan (1980) verschiedene Aktivitäten, die sich auf die Beobachtung von Systemzuständen oder Resultaten von Maßnahmen beziehen können, während es bei Goodman(1968) die steuernde Überwachung von Ressourceneinsatz bedeutet.

Es treten zudem häufig in den Arbeiten Sammelbezeichnungen auf, die ebenfalls die eindeutige Zuordnung im Rahmen des Kategoriensystems erschweren, wie zum Beispiel Planning, Coordinating als eine einzige Führungsdimension oder ebenso Decision making, Problem solving bei Luthans und Lookwood (1982). Auch diesen für die eindeutige Zuordnung problematischen Fall erleichtert die enge Kontextexplikation. Außerdem zeigt sich vor allem bei den sehr differenzierten Instrumenten, dass verschiedene Komponenten in dieselbe Teilkategorie klassifiziert werden. So wurden auch Literaturbefunde, die sich eigentlich auf den interpersonellen Bereich beziehen, teilweise wegen ihrer Aufgabeninhalte in problemlösebezogene Kategorien eingeordnet.

Nicht alle Haupt- und Teilkategorien kamen gleichermaßen häufig zum Einsatz, einige sogar überhaupt nicht. Dennoch sollen sie für den Hauptlauf beibehalten werden, da auch interessiert, inwieweit das Rahmenmodell mit dem Schwerpunkt auf Anforderungen an komplexe Problemlöser nach Dörner (1986) durch die Befunde und Theorien zu Führungsaufgaben abgedeckt wird. Weniger geht es darum, dieses Modell auf der Basis der ausgewähl-

ten Führungsliteratur zu modifizieren. Eine Erweiterung wird also nicht als nötig betrachtet.

8.4.2 Hauptkategorien

Die Ergebnisse werden insgesamt und getrennt nach der Art der Arbeit anhand deskriptiver Statistiken dargestellt. Weitere inferenzstatistische oder multivariate Auswertungsverfahren wären für die Fragestellung sicherlich interessant und sinnvoll, verbieten sich aber angesichts der Stichprobengröße und des Skalenniveaus. Für die Übersicht über die fünf Bereiche des Kategoriensystems wurden zunächst über alle Arbeiten hinweg die Kodierungshäufigkeiten ausgezählt und addiert, ohne nach den drei Arten von Führungsliteratur zu trennen, wobei außerdem die Nennungen der Teil- und der Hauptkategorien additiv zusammengefasst wurden. Im nächsten Schritt wurde die mittlere Kodierhäufigkeit für alle Arbeiten sowie für die drei Literaturkategorien getrennt bestimmt.

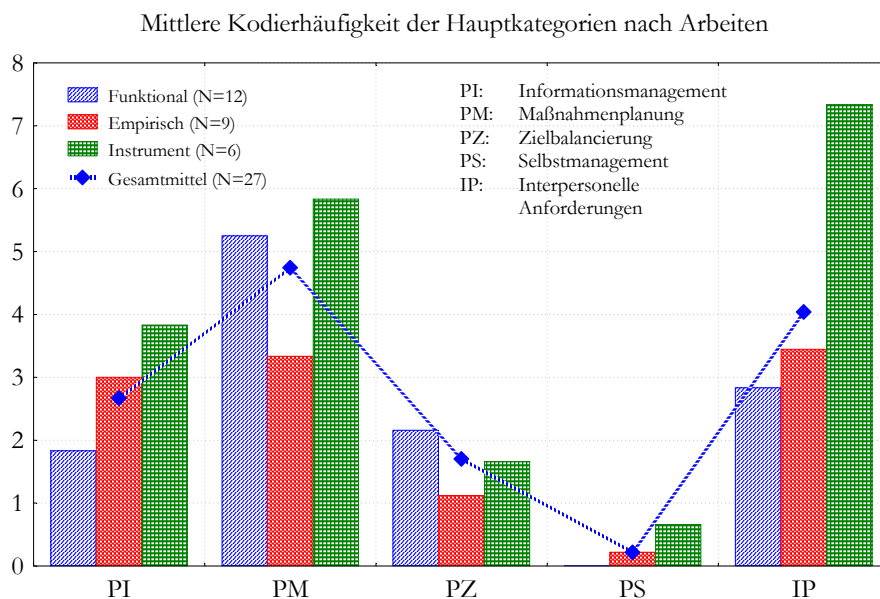


Abbildung 8.3: Mittlere Kodierhäufigkeiten der Hauptkategorien nach Arbeiten.

Das eingesetzte Kodierschema hatte fünf Hauptkategorien mit jeweils sechs oder drei Teilkategorien. Davon deckten vier Hauptkategorien die vier

Anforderungsdimensionen an die Löser komplexer Probleme aus dem Modell von Dörner (1986) ab. Dabei handelt es sich um Informationsmanagement, Zielbildung Maßnahmenplanung und Selbstmanagement. Die fünfte Hauptkategorie umfasste einige Aufgaben, die sich auf interpersonelle Anforderungen beziehen.

Zunächst erfolgte eine Auszählung der Kodierungen für jeden Abschnitt des Kategoriensystems insgesamt, also gemeinsam für Haupt- und Teilkategorien, um festzustellen, wie weit die fünf Bereiche des Kategoriensystem überhaupt genutzt wurden. Insgesamt ließen sich die Ergebnisse der Führungsliteratur gut in die Kategorien zum Problemlösen einordnen, sieht man von dem Bereich des Selbstmanagements ab. Dabei sind die am häufigsten kodierten Kategorien die Maßnahmenplanung und das Informationsmanagement, gefolgt vom Bereich Zielbildung. Der soziale Bereich wurde insgesamt ebenfalls sehr häufig abgedeckt.

Die unterschiedlichen Häufigkeiten hängen möglicherweise auch damit zusammen, dass die fünf Bereiche unterschiedlich viele Teilkategorien haben. So gliedern sich Informationsmanagement, Maßnahmenplanung und der soziale Bereiche in jeweils sechs Teilkategorien auf, Zielbildung und Selbstmanagement enthalten nur drei Teilkategorien. Allerdings konnte immer auch auf der Ebene der Hauptkategorie kodiert werden. Zur besseren Vergleichbarkeit wurde die Kodierhäufigkeit der gesamten Führungsliteratur und der drei Literaturgattungen gemittelt, wobei die unterschiedliche Anzahl von Teilkategorien ebenfalls durch entsprechende kompensative Gewichtung bei der Mittelung berücksichtigt wurde. Dabei weisen Instrumente bei fast allen Kategorien die größte Kodierungshäufigkeit auf, was darauf zurückzuführen ist, dass sie die differenzierteste und umfangreichste Darstellung der Komponenten von Führungstätigkeit beinhalten, und damit wohl zum Zwecke der Operationalisierung, die geringste Abstraktionsebene mit ihren Skalen besitzen.

Relativ gesehen, wurde bei den Instrumenten der interpersonelle Bereich am häufigsten kodiert, gefolgt von Maßnahmenplanung und Informationsmanagement. Die funktionalen Verfahren hatten ihren Schwerpunkt bei den Aspekten der Maßnahmenplanung, die am häufigsten kodiert wurden, gefolgt vom sozialen Bereich und Informationsmanagement. Selbstmanagement wurde gar nicht erwähnt. Bei den empirischen Studien unterschied sich die Kodierhäufigkeit von Informationsmanagement, Maßnahmenplanung und sozialem Bereich nur geringfügig. Es scheint sich also bereits bei der Verteilung der Verfahren auf die fünf Bereiche ein unterschiedliches Muster von Kodierhäufigkeiten zu ergeben, unterscheidet man nach der Art der Arbeiten.

8.4.3 Teilkategorien

Informationsmanagement umfasst laut Dörner (1986) die Aktivitäten zur Beschaffung und sinnvollen Verwertung notwendiger Informationen über ein Problem, über Zusammenhänge der Situation, über Maßnahmen und deren Wirkung sowie über den Zustand kritischer Variablen. Darin enthalten sind Aspekte der Orientierung und Kontrolle sowie Strategien der Informationsbeschaffung. In diesem Sinne enthält der Abschnitt Informationsmanagement sechs Teilkategorien.

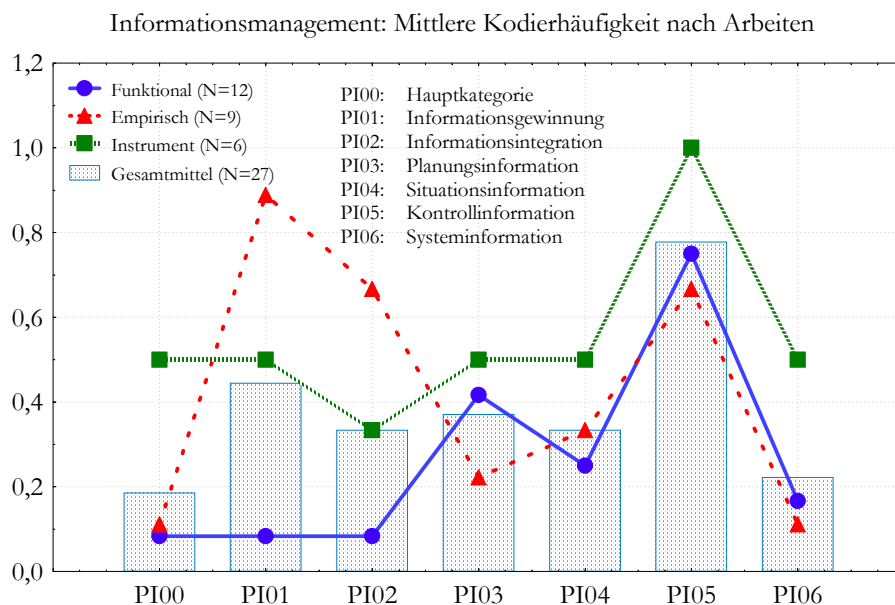


Abbildung 8.4: Mittlere Kodierhäufigkeiten Informationsmanagement nach Arbeiten.

Analog zum Vorgehen bei den Gesamtergebnissen wurde zunächst die gesamte Häufigkeit der Kodierungen für die 27 Arbeiten ausgezählt, allerdings hier getrennt nach den Nennungen in den Haupt- und den Teilkategorien des Abschnitts Informationsmanagement. Dabei zeigt sich, dass alle Teilkategorien genutzt wurden und damit keine davon überflüssig war. Ebenso fanden, wenn auch am seltensten, Kodierungen nur auf der Ebene der Hauptkategorie statt, da die Analyseeinheiten anhand der Ausführungen im Material nicht genauer spezifiziert werden konnten. Am häufigsten wurde jedoch die Teilkategorie Kontrollinformation belegt, wobei es darum geht, die Resultate

eigener Handlungen oder Maßnahmen zu überprüfen.

Mittelt man diese Häufigkeiten und unterteilt sie nach Arbeiten im Vergleich zum Gesamtmittel, ergeben sich unterschiedliche Profile der jeweiligen Literaturbeiträge. In empirischen Arbeiten werden die Kategorien Informationsbeschaffung und Informationsintegration am häufigsten erwähnt, während diese bei den funktionalen Arbeiten kaum vorkommen. Die Kategorie Kontrollinformation lässt sich dagegen am häufigsten bei den Instrumenten kodieren.

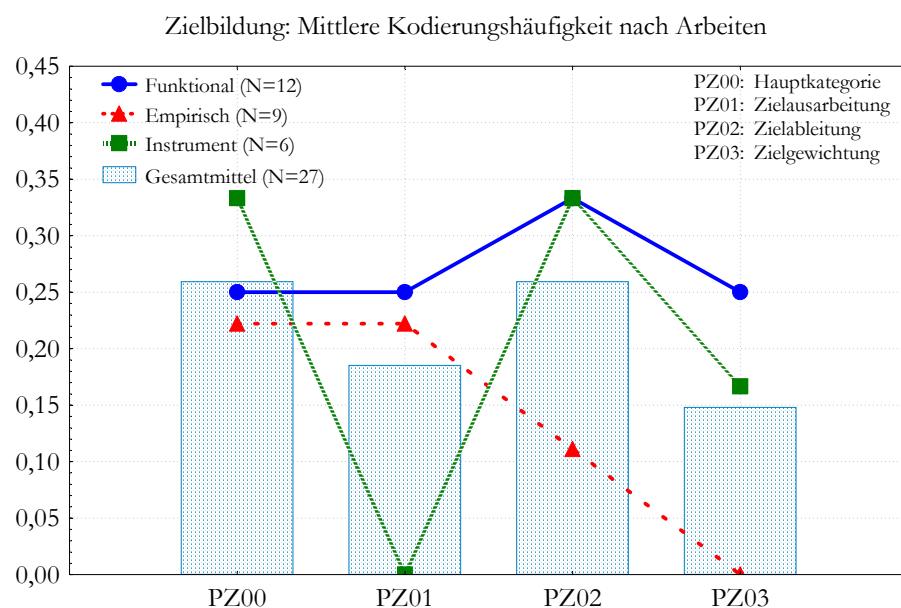


Abbildung 8.5: Mittlere Kodierhäufigkeiten Zielbildung nach Arbeiten.

Eine weitere Anforderung im Rahmen des Problemlösemodells ist die Zielausarbeitung, wenn Ziele vage oder global vorgegeben sind. Daraus müssen Teilziele abgeleitet und konkretisiert werden, die unter Umständen im Widerspruch zueinander stehen, weshalb des Weiteren eine entsprechende Gewichtung der Ziele, eine Zielbalancierung erforderlich wird, die bis zum Verzicht auf Teilziele führen kann. Dies wurde im Kategoriensystem mit dem Abschnitt Zielbildung umgesetzt, der drei Teilkategorien enthält.

Auch wenn die Zielbildung insgesamt einen eher weniger genutzten Abschnitt des Kategoriensystems darstellt, so verteilen sich die Kodierungshäufigkeiten für alle Arbeiten zusammen kontinuierlich auf die Hauptkategorie

und die Teilkategorien. Am häufigsten ließen sich die Analyseeinheiten auf der Ebene der Hauptkategorie sowie der Teilkategorie Zielableitung einordnen, gefolgt von Zielausarbeitung und Zielgewichtung.

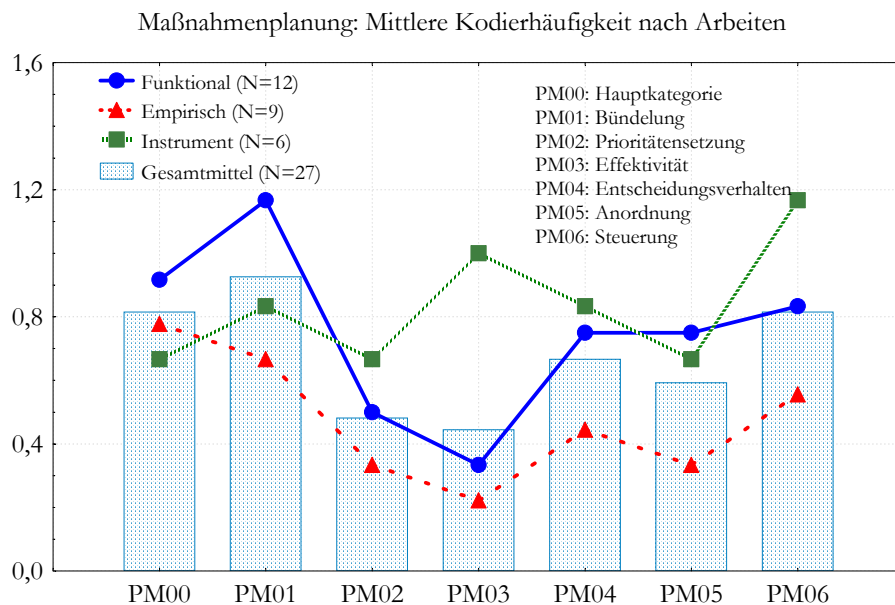


Abbildung 8.6: Mittlere Kodierhäufigkeiten Maßnahmenplanung nach Arbeiten.

In der Regel enthält die Maßnahmenplanung ein Bündel von Einzelmaßnahmen, die sich auf verschiedene Eingriffsvariablen bezüglich des Zieles beziehen. Hierbei geht es meist auch nicht allein um die Anordnung einer punktuellen Maßnahme oder eines solchen Bündels, sondern die Maßnahmen müssen in eine Reihenfolge aneinander gebracht werden, wobei spätere Maßnahmen auf die Resultate vorhergegangener Maßnahmen aufbauen.

Kategorien aus dem Abschnitt Maßnahmenplanung wurden nicht nur insgesamt am häufigsten kodiert, sondern auch seine Teilkategorien ließen sich für die Einordnung des Materials offensichtlich alle gut gebrauchen. Dabei erschienen am häufigsten die Bündelung von Maßnahmen, die Steuerung und die Hauptkategorie. Etwas weniger häufig traten Entscheidungsverhalten und Anordnung auf, am seltensten Prioritätensetzung und Effizienz.

Die Eigenschaften, die eine Situation als komplexes Problem charakterisieren, führen zu Zeitdruck, Unsicherheit über die Wirkung der angeordneten Maßnahmen und damit zu Stress. Eventuelle Misserfolge bei der Steuerung

erzeugen Frustration. Möglicherweise muss man Maßnahmen anordnen, die dem eigenen Wertesystem widersprechen, wie Lohnkürzungen oder Entlassungen in einem Betrieb. Deshalb ist emotionale Belastbarkeit im Umgang mit diesen Faktoren erforderlich. Dörner (1986) bezeichnet diese weniger kognitive als emotionale Fähigkeit von Personen, die eigenen Fähigkeiten und Gewohnheiten an die Situation anzupassen, als Selbstmanagement, die im Sinne einer Anforderung nach den eher knappen Ausführungen dort im Rahmen des Kategoriensystems mit drei Teilkategorien umgesetzt wurde.

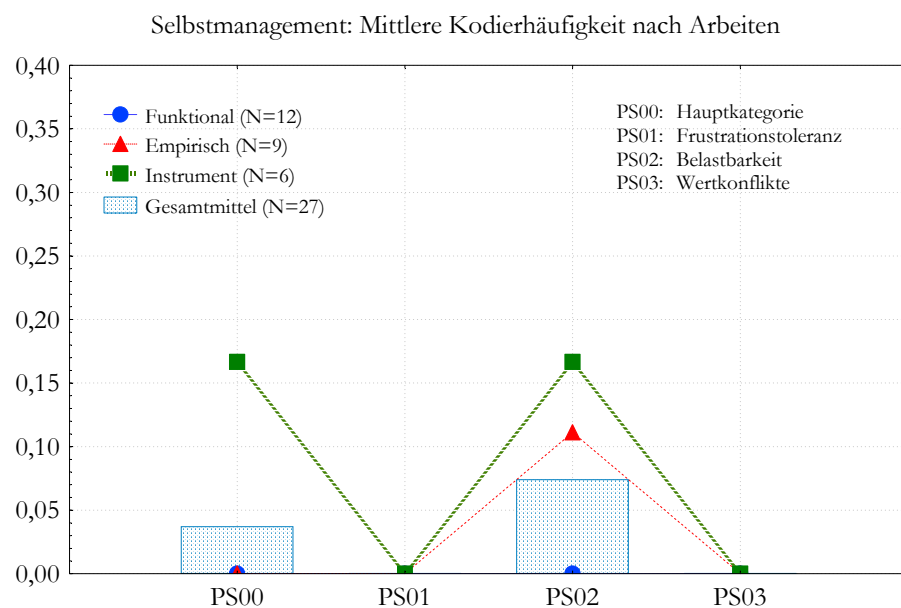


Abbildung 8.7: Mittlere Kodierhäufigkeiten Selbstmanagement nach Arbeiten.

Obwohl Dörner (1976) die Relevanz von Selbstmanagement in eigenen Untersuchungen veranschaulichen kann und außerdem das Konzept durch hohe Plausibilität besticht, lässt sich kaum eine der Kategorien für die Kodierung der Führungsaspekte nutzen, da Selbstmanagement in dieser Form kaum thematisiert wird. Von den Teilkategorien wurde Belastbarkeit zweimal genannt, außerdem konnte einmal die Hauptkategorie kodiert werden. Die beiden anderen Teilkategorien kamen überhaupt nicht zum Einsatz. Das mag daran liegen, dass Selbstmanagement nur eine mittelbare Anforderung an Problemlöser darstellt, bedingt durch die Eigenschaften der Situation, und damit nicht explizit benannt wird, oder an einer zu stark technischen Per-

spektive der einbezogenen Führungsliteratur. Angesichts der insgesamt drei Nennungen erübrigen sich weitere Diskussionen über die Feststellung hinaus, dass die Haupt- und Teilkategorien Selbstmanagement für das Material keine fruchtbaren Kategorisierungsperspektiven liefern.

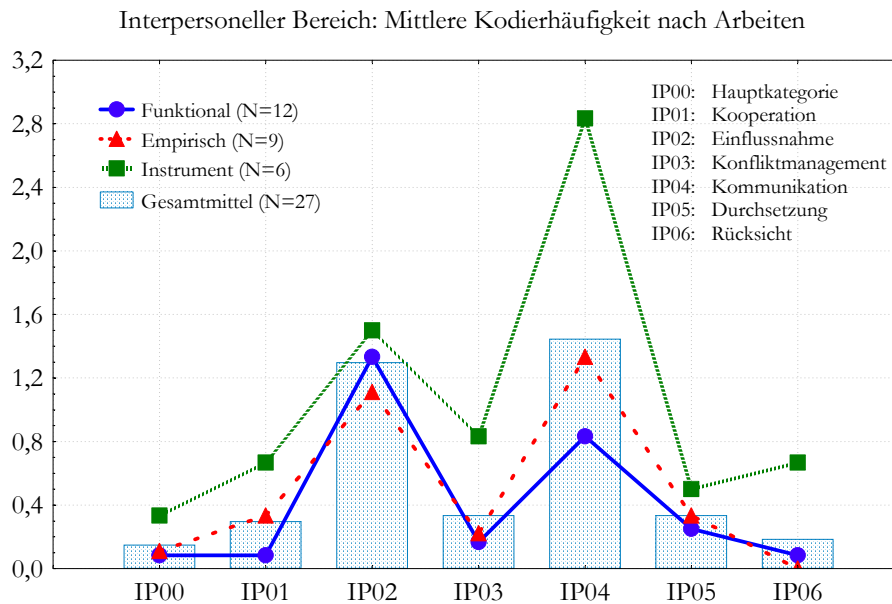


Abbildung 8.8: Mittlere Kodierhäufigkeiten interpersoneller Bereich nach Arbeiten.

Der mit interpersoneller Bereich bezeichnete Abschnitt des Kategoriensystems stellt eine Sammelkategorie für Anforderungen an Führungskräfte dar, die im alltäglichen ebenso wie im wissenschaftlichen Sprachgebrauch vage als soziale Kompetenz bezeichnet, aber nicht klarer operationalisiert werden konnten. Eine denkbare Basis wäre die Definition interpersonaler Kompetenz von Weinstein (1969), die hier in der Fähigkeit besteht, Aufgaben im Umgang mit anderen Personen zu bewältigen und das Verhalten anderer beeinflussen zu können. Ergänzend können die Ansätze zur Selbstbehauptung in der klinischen Psychologie genutzt werden. Diese besteht darin, in Konflikten adäquat zu reagieren, seine eigenen Bedürfnisse zu artikulieren und durchzusetzen, ohne dabei andere zu beeinträchtigen. Solche und weitere Überlegungen wurden zu sechs eher allgemein gehaltenen Teilkategorien verdichtet.

Insgesamt wurde die Hauptkategorie mit ihren Teilkategorien häufig kodiert. Allerdings verteilte sich die Kodierungshäufigkeit sehr ungleichmäßig über die Haupt- und Teilkategorien des interpersonellen Bereichs, wobei deutlich am intensivsten die Kommunikation und die Einflussnahme genutzt wurden, wenn auch die anderen Teilkategorien sowie die Hauptkategorie gelegentlich zum Einsatz kamen. Vielleicht liegt dies daran, dass Kommunikation als häufigste Nennung sehr allgemein definiert ist, und soziale Einflussnahme in den meisten Ansätzen ein charakterisierender Bestandteil der Definition von Führung ist.

8.5 Fazit

Die literaturgestützten Anforderungsanalyse sollte einen Beitrag zur Evaluation der Managementdiagnostik leisten, indem sie ein Forschungsmodell für eine normative Outcome-Evaluation erarbeitet. Insgesamt scheint es möglich zu sein, eine Stichprobe der ausgewählten Führungsliteratur auf das Kategoriensystem mit dem Schwerpunkt auf Problemlösen anzuwenden, da sich alle Befunde kodieren ließen. Ob dies auch durch andere Kodierer in der gleichen Weise gelingt, müsste noch untersucht werden. Insbesondere die Bereiche Informationsmanagement und Maßnahmenplanung können die Aspekte von Führungstätigkeit gut abbilden. Die Sammelkategorie interpersoneller Bereich wird zwar oft genutzt, allerdings hauptsächlich mit den Teilkategorien Kommunikation und Einflussnahme.

Die Kategorie Selbstmanagement wurde bei insgesamt 27 Studien nur dreimal kodiert. Die Frage ist, ob dies ein Defizit des Problemlösemodells darstellt, das hiermit eine irrelevante Anforderungsdimension abbildet, oder ein Defizit der Führungsliteratur, die dies nicht berücksichtigt. Dies ließe sich dadurch klären, indem bei Führungskräften erfasst wird, wie gut sie diese Anforderung bewältigen, und ob dies mit dem Berufserfolg entsprechend korreliert. Außerdem könnte im Rahmen einer eigenen Anforderungsanalyse untersucht werden, ob Aspekte des Selbstmanagements tatsächlich eine relevante Anforderung bei Führungstätigkeiten darstellen. Abgesehen von den inhaltlichen Aspekten wurde mit der Anwendung der qualitativen Inhaltsanalyse als Methode zur Literatursynthese ein methodischer Zugang entwickelt, der die bisherigen metaanalytischen Strategien ergänzen kann. So lassen sich begrifflich heterogene, konzeptuelle und interdisziplinäre Arbeiten integrieren.

Kapitel 9

Diskussion

Evaluation der Managementdiagnostik

Unter Evaluation wurde in vielen eignungsdiagnostischen Studien häufig nur Kriteriumsvalidierung verstanden. Jedoch nicht nur hier greift ein solcher Ansatz zu kurz, der eine Maßnahme alleine vom Kriterium her induktiv bewertet, da auf diese Weise viele wichtige Aspekte, gerade in Anwendungssituationen, übersehen werden. Evaluationsstudien der Diagnostik können weitere Strategien zur Gewinnung diagnostischer Gütekriterien und zur Evaluierung des Urteilsprozesses beinhalten. Zudem sind im Anwendungskontext weitere Aspekte wichtig, insbesondere eine Balancierung der Anwenderperspektive und inhaltlicher Theorien, um die Akzeptanz und Kommunizierbarkeit der Resultate zu optimieren. Deshalb müssen für eine solche Evaluationsforschung umfassendere Modelle gefunden werden. Gleichzeitig sollten sie spezifischer auf diese Anwendung zugeschnitten sein.

Deshalb wurde der allgemeine Ansatz der „theory driven evaluations“ (Chen, 1990) angewendet, um verschiedene Aspekte der Managementdiagnostik zu integrieren und zudem die Perspektive der Auftraggeber zu berücksichtigen. Methodische Strategien im gesamten diagnostischen Prozess werden in die Evaluationstypologie auf einer allgemeinen Ebene eingeordnet und in den nachfolgenden Studien beispielhaft veranschaulicht.

Die kausativen Evaluationstypen, die sich hauptsächlich auf verschiedene Aspekte des erfassten Kriteriums beziehen, bilden eher vertraute Aspekte der Eignungsdiagnostik und damit verbundener Evaluations- oder Validierungskonzepte ab, weshalb in den vorgestellten Studien normative Evaluationstypen im Vordergrund stehen. Die normativen Typen stellen eine Erweiterung der Input-Output-Sichtweisen dar, welche in vielen Evaluationsstudien noch vorherrscht, wobei sie dafür sensibilisieren, auch die Umsetzung der Maßnahme, Urteilsprozesse und Zieldefinitionen zu bewerten und zu begleiten.

Aus der Sicht der Diagnostik werden einerseits bekannte Strategien zur Güteüberprüfung neu geordnet, darüber hinaus werden der Umgang mit dem Stakeholder und die damit verbundenen Aspekte stärker betont. Sofern also eine Evaluation der Managementdiagnostik gewünscht wird, oder die Managementdiagnostik den Anspruch erhebt, sich auf das Feld der Evaluationsforschung zu begeben, wie viele andere psychologischen Disziplinen es bereits getan haben, könnte dies eine Möglichkeit sein, den Evaluationsprozess zu strukturieren und zu erweitern, sowie fachspezifisch zu füllen.

Auswertung der Stellenanzeigen

Eine Auswertung von Stellenanzeigen bei 83 Unternehmen bildete als Beitrag zur Ziel-Evaluation fachliche und nichtfachliche Anforderungsprofile aus der Perspektive der Praxis ab, wie sie gegenüber Bewerbern kommuniziert werden. Diese erscheinen relativ allgemein, und weisen nur wenig inhaltliche Bezüge zu psychologischen Eigenschaftskonstrukten auf. Dennoch bringen sie implizite Eignungstheorien zum Ausdruck und sind von hoher Relevanz, weil diese die Personalarbeit prägen. Als solche haben sie großen Einfluss auf Zielbildungsprozesse für Evaluationsstudien, natürlich auch generell für die Managementdiagnostik unabhängig von jeglicher Evaluation.

Besonders interessant sind die außerfachlichen Qualifikationen. Am häufigsten wurden bei über 50 Prozent der Unternehmen jeweils Flexibilität und Engagement genannt, was auf kognitive und motivationale Merkmale schließen lässt, während nur ein Drittel soziale Kompetenz angaben. Jedoch erschienen als weiteres spezifische kognitive Merkmal analytische Fähigkeiten am seltensten. Diese Befunde basieren auf Stellenanzeigen, die im außerfachlichen Bereich häufig eher allgemein gehalten sind. Es wäre vor diesem Hintergrund interessant, die zuständigen Personalfachleute selbst zu ihren impliziten Eignungstheorien bei Führungskräften zu befragen, um ein differenzierteres Bild zu bekommen.

Daneben gibt die Studie einen allgemeinen Eindruck über verwendete Auswahlverfahren. Dennoch wurde zur Verbreitung verschiedener Auswahlverfahren in jüngerer Zeit erstaunlich wenig in deutschen Fachblättern publiziert. Auch hierfür wäre es wünschenswert, noch mehr und genauere Daten zu erhalten. Bekanntlich führen Unternehmensberatungen auch in diesem Bereich Benchmarks und Best-Practice-Studien durch, die meist nicht in psychologischen oder betriebswirtschaftlichen Fachzeitschriften veröffentlicht werden, wie etwa die Personalberatung Kienbaum (Schmidt-Klingenberg, 2002).

Kompetenzmodell im Konzern

Die Dokumentation einer unternehmensinternen Entwicklung eines Kompetenzmodells für obere Führungskräfte zeigt eine Möglichkeit, die Praxisperspektive mit einer wissenschaftlichen Perspektive zu verbinden und betont die Notwendigkeit einer kontinuierlichen Abstimmung zwischen Evaluatoren und Auftraggebern oder Stakeholdern, damit die Ergebnisse für alle Beteiligten weiterverwendet werden können. In einem mehrstufigen Prozess wurden neben dem wissenschaftlich begründeten Modell mehrere Perspektiven integriert. Ausgehend von dem Problemlösemodell wurde dieses erweitert um die besondere Unternehmenssituation, um die Erwartungen erfolgreicher Stelleninhaber sowie verschiedener Leiter der Personalabteilungen im Konzern. Diese Definition eines Anforderungsmodells für die anschließend geplante interne Personalentwicklung stellt eine *normative outcome evaluation* dar, wobei als Methode nominale Gruppen eingesetzt wurden.

Ergänzend zur nominalen Gruppenmethode hätten quantitative Verfahren zur Strukturierung von Entscheidungsprozessen und Diagnose von Zielbeziehungen eingesetzt werden konnten, wie sie in der DELPHI-Methode oder der multiattributiven Nutzenanalyse enthalten sind. Dennoch zeigte sich hier, dass ein Beitrag aus wissenschaftlicher Sicht nicht generell in der Praxis abgelehnt wird, sondern die wissenschaftliche Begleitung neben methodischer Unterstützung auch inhaltliche Beiträge leisten kann.

Metaanalyse Persönlichkeit

Die Metaanalyse schließlich bildet die Bedeutsamkeit wissenschaftlicher Persönlichkeitskonstrukte als Bestandteil von Anforderungsprofilen für den Berufserfolg bei verschiedenen Stichproben von Führungskräften ab und gibt erste Hinweise darauf, welche davon für weitere diagnostische Maßnahmen wertvoll sein könnten. Es wurden 16 neuere Studien mit insgesamt 18 Untersuchungsstichproben ausgewertet. Für jede Big-Five-Dimension lagen maximal 16 (Extraversion) und mindestens 8 (Offenheit) Korrelationskennwerte vor. Die Gesamtstichprobe umfasste 10398 Teilnehmer. Ausgewertet wurden die Korrelationskoeffizienten.

Hieraus resultierten für den Zusammenhang von Neurotizismus und Extraversion mittlere Effekte. Es ist also eine hohe Extraversion sowie eine hohe emotionale Stabilität günstig für Führungserfolg, während die anderen drei Persönlichkeitsdimension Offenheit für neue Erfahrungen, Gewissenhaftigkeit und Verträglichkeit keine Rolle spielen. Mit den Ergebnissen für Extraversion und Neurotizismus entsprechen diese Ergebnisse den Befunden für den Zusammenhang mit Berufserfolg bei den meisten Berufsgruppen, während

Gewissenhaftigkeit im Unterschied dazu bei Führungskräften eine kleinere Rolle spielt. Es können also nicht alle Befunde von anderen Berufsgruppen auf Führungskräfte übertragen werden. Bei einer größeren Stichprobe wäre es interessant, die Auswertung nach unterschiedlichen Erfolgskriterien aufzuteilen sowie nach Stichproben, die aus unterschiedlichen Branchen und Führungsebenen stammen.

Treatment-Evaluation

Die Personalabteilung eines Technologie-Konzerns wünschte post hoc eine Evaluationsstudie ihrer internen Managementdiagnostik bei oberen Führungskräften, von der Teilnehmerdaten vorlagen. Es war Aufgabe der wissenschaftlichen Begleitung, die Entwicklung von Evaluationskriterien durch die Auftraggeber zu unterstützen, um in einer zweiten, quantitativen Phase statistische Auswertungsstrategien vorzuschlagen und umzusetzen.

Neben den Auswertungsstrategien im Sinne einer normativen Treatment-Evaluation wurden die Daten für eine kausative Impact-Evaluation im Sinne des wissenschaftlichen und des betrieblichen Anforderungsmodells modelliert. Im Mittelpunkt stand die Modellierung und Begleitung der Urteilsprozesse auf verschiedenen Ebenen. Da der Datensatz die Voraussetzungen dafür erfüllte, ließen sich diese Varianten teilweise mit multivariaten statistischen Verfahren abbilden und prüfen. Weitere Zielkriterien wurden ebenfalls statistisch modelliert und geprüft. So reproduzierte die faktorenanalytische Multitrait-Multimethod-Validierung bekannte Ergebnisse aus der Assessment-Center-Forschung, während die Objektivität der Urteiler anhand der Auswertungen gegeben war.

Von unternehmenspolitischer Relevanz stellte sich die Modellierung der Urteilsprozesse zwischen den beteiligten Beratungsunternehmen dar, die empirisch anders ausfiel als von den Auftraggebern gewünscht. An diesen und andere Stellen konnte die wissenschaftliche Begleitung selbst Vorschläge zur Verbesserung der Maßnahme einbringen und umsetzen. Im Unterschied zu Untersuchungen, die von Anfang an in der Hand von Wissenschaftlern liegen, bestand hier die Herausforderung darin, dass der Evaluator wenig Einfluss darauf hatte, welche Daten in welcher Form zur Verfügung gestellt wurden. So waren die Evaluationsdaten bereits stark aggregiert. Solche Voraussetzungen machten dieses Projekt gleichzeitig interessant, da sie unmittelbar die Praxisbedingungen widerspiegeln und das Projekt als solches eine hohe praktische Relevanz aufweist.

Impact-Evaluation

Ein eigenes Anforderungsmodell zum Problemlösen und zu interpersonellen Fertigkeiten bildete die Grundlage für eine diagnostische Untersuchung an 68 Nachwuchsführungskräften und das Rahmenkonzept zur Ableitung der Untersuchungshypothesen auf Prädiktor- und Kriterienseite. Von besonderem Interesse war ein computergestütztes komplexes Szenario. Weitere Maße aus dem beruflichen und dem Persönlichkeitsbereich ergänzten die Prädiktoren. Fragestellung war, wie weit die Problemlöseleistung sowie die anderen Prädiktoren beruflichen Erfolg vorhersagen können, es wurde also eine konkurrenthe Kriteriumsvalidierung vorgenommen. Als berufliches Leistungskriterium wurden die Ergebnisse einer Arbeitsprobe modelliert. Außerdem sollte die Beziehung zwischen Problemlöseskalen und Maßen aus dem interpersonellen und dem Persönlichkeitsbereich geklärt werden.

Selbst wenn man die methodischen Einschränkungen mit einbezieht, die sich in dieser Praxissituation ergaben, lässt sich diese Untersuchung durch die modellkonformen Korrelationsmuster als erfolgreich bewerten. Dadurch, dass die Prädiktoren und das Kriterium im Sinne des Problemlösemodells modelliert und aggregiert wurden, konnten aus den vielen denkbaren Auswertungen wenige, relevante Analysen ausgewählt werden. Diese erfuhren eine weitere Einschränkung und damit einen Gewinn an Informationsgehalt, indem ebenfalls aus dem Problemlösemodell gezielte Erwartungen abgeleitet werden konnten, an welcher Stelle und in welcher Form Zusammenhänge auftreten sollten.

Dennoch wäre es im Sinne der Fragestellung günstiger gewesen, wenn das diagnostische Instrumentarium sowie die Auswertung direkt und durchgängig auf der Basis des hier erarbeiteten Modells implementiert wäre, statt die Daten, die eigentlich auf dem betrieblichen Anforderungsmodell beruhen, entsprechend neu zu ordnen. Zudem waren die Ratings der Vorgesetztenbeurteilung unbrauchbar, was eine umfassendere Auswertung unter Einbeziehung interpersoneller Anforderungen, wie sie ursprünglich gedacht war, verhinderte. Im Falle der Selbstreflexion und der Hausaufgabe ließen sich Urteilsverzerrungen durch die modellgeleitete Auswertungsstrategie teilweise kompensieren.

Das computergestützte Szenario hatte dagegen nicht diese Probleme und war gleichzeitig ökonomischer. Es schien also selbst im Kontext des außerwissenschaftlichen Einsatzes sehr robust, da es die Durchführungs- und Auswertungsobjektivität optimiert. Auch die Teilnehmer kamen gut damit zurecht. Abgesehen vom computergestützten Szenario, das explizit auf der Grundlage des hier verwendeten Problemlösemodells entwickelt wurde, lassen sich auch für die weiteren hier verwendeten Verfahren Alternativen denken, ins-

besondere für die Indikatoren beruflicher Leistung. Angesichts dessen, dass es aktuell noch wenig Befunde über den Zusammenhang von computergestützten Szenarios mit Berufserfolgskriterien gibt, kann diese Untersuchung als weiterer Beitrag gewertet werden. Besonders deutlich fiel der hypothesenkonforme Zusammenhang zwischen Problemlösemaßen des Szenarios und leistungsbezogenen Berufserfolgskriterien aus.

Es lässt sich festhalten, dass das verwendete Problemlösemodell als Hintergrund für die Aggregation der Daten und außerdem für die Formulierung von Hypothesen offensichtlich eine brauchbare Grundlage lieferte. Die erzielten Korrelationen als Koeffizienten für die konkurrente Kriteriumsvalidität entsprachen mittleren Effektgrößen. Multivariate Analyseverfahren wären angesichts der Fragestellung interessanter, aber dafür war die Stichprobe zu klein.

Die Untersuchung fand in einer echten Auswahl-situation und außerdem an einer Stichprobe von Nachwuchsführungskräften statt, also genau in der intendierten Anwendungssituation. Neben der praktischen Relevanz bedeutet das, dass man davon ausgehen kann, dass die Teilnehmer motiviert waren und vor allem die Ergebnisse der eher kognitiv leistungsorientierten Verfahren wie FSYS und die Arbeitsprobe repräsentativ für die obere Leistungsgrenze der Teilnehmer waren. Ob und wie weit bei den anderen Verfahren Simulation oder Dissimulation stattgefunden hatte, kann im Nachhinein schwer festgestellt werden. In den Persönlichkeitstests waren Lügen- oder Kontrollskalen enthalten, diese konnten jedoch keinen Beitrag zur Aufklärung der Ergebnisse in den anderen Verfahren leisten.

Literaturgestützte Anforderungsanalyse

Seit einigen Jahren verwendet die psychologische Anwendungsforschung die qualitative Inhaltsanalyse, um sprachliches Material auszuwerten. Diese Methode wurde in dieser Arbeit für die Literatursynthese adaptiert, um die Befunde zu Führungsaufgaben vor dem Hintergrund eines Problemlösemodells einzuordnen und gleichzeitig ein Forschungsmodell vorzustellen, das Beiträge zur Managementdiagnostik und deren Evaluation liefern könnte. Damit wurde die Frage bearbeitet, ob Führungs- oder Managementaufgaben vergleichbare Anforderungen wie komplexe Probleme stellen.

Bisher gab es kaum Methoden, um heterogene Literatur zu integrieren. In eine Metaanalyse lassen sich nur solche Arbeiten einbeziehen, welche dieselben Begriffe verwenden, ein vergleichbares Design haben und deren Ergebnisse in Form von überführbaren statistischen Kennwerten vorliegen. Die Methode zur Literaturanalyse, die hier auf der Grundlage der qualitativen Inhaltsanalyse nach Mayring (1997) entwickelt wurde, war ein Versuch, diese

Lücke in der Auswertung von heterogenen Literaturbefunden zu schließen, ohne dabei den Anspruch auf Systematik aufzugeben.

Weil die Befunde der Führungsliteratur nach inhaltlichen Gesichtspunkten geordnet werden sollen, war eine inhaltliche Strukturierung die gewählte Durchführungsvariante der qualitativen Inhaltsanalyse. Schließlich wurde dargestellt, wie sich das zu analysierende Material aufbereiten lässt. Wichtig ist dabei, dass die üblicherweise als Führungsdimensionen dargestellten Ergebnisse durch die definitorischen Beschreibungen so erweitert werden, dass sie sich in das Kategoriensystem einordnen lassen. Damit wurde das Rahmenmodell der inhaltlichen Strukturierung, dessen Kern das Kategoriensystem bildet, durch eine weitere inhaltsanalytische Technik ergänzt, nämlich durch eine enge Kontextexplikation. Dies ermöglicht, über eine reine Auszählung der oft irreführenden Oberbegriffe für Führungsdimensionen hinauszugehen.

Diese Ergebnisse lassen sich unter zwei Gesichtspunkten betrachten. Hier wurde das Kategoriensystem weitgehend vollständig ausgenutzt, und es blieben keine Kodiereinheiten übrig, die sich nicht einordnen ließen. Insgesamt scheint es möglich zu sein, nach der erarbeiteten Methodik eine Stichprobe der ausgewählten Führungsliteratur in das Kategoriensystem mit dem Schwerpunkt Problemlösen einzuordnen, da sich alle Befunde kodieren ließen. Insbesondere die Bereiche Informationsmanagement und Maßnahmenplanung konnten die Aspekte von Führungstätigkeit gut abbilden. Die Sammelkategorie interpersoneller Bereich wurde zwar oft genutzt, allerdings hauptsächlich mit den Teilkategorien Kommunikation und Einflussnahme, die sehr allgemein sind.

Auffällig ist das Ergebnis zur Kategorie Selbstmanagement. Sie wurde bei insgesamt 27 Studien nur dreimal kodiert. Die Frage ist, ob dies ein Defizit des Problemlösemodells darstellt, das hiermit eine irrelevante Anforderungsdimension abbildet, oder ein Defizit der Führungsliteratur, die dies nicht berücksichtigt. Dies ließe sich nur dadurch beantworten, indem bei Führungskräften selbst erfasst wird, wie gut sie diese Anforderung bewältigen, und ob dies mit dem Berufserfolg entsprechend korreliert. Außerdem könnte im Rahmen einer eigenen Anforderungsanalyse untersucht werden, ob Aspekte des Selbstmanagements tatsächlich eine relevante Anforderung bei Führungstätigkeiten darstellen.

Das bedeutet noch nicht, dass dies die einzige Möglichkeit ist, die Führungsliteratur in allen Bedeutungsaspekten erschöpfend zu systematisieren, zumindest ist dies ein Hinweis, dass eine solche Strukturierung eine denkbare Möglichkeit darstellt. Ob dies auch durch andere Kodierer in der gleichen Weise gelingt, müsste noch untersucht werden. Zumindest stellt dieses Vorgehen einen Versuch dar, führungsrelevante Anforderungen systematisch zu erschließen, der soweit sich hier feststellen lässt, erfolgreich war. Zur breiteren

Erfassung der Stakeholder-Perspektive wäre es sicherlich interessant, auch populärwissenschaftliche Literatur und Ratgeberbände ebenfalls inhaltsanalytisch auszuwerten und den Ergebnissen der wissenschaftlichen Literatursynthese gegenüberzustellen. Möglicherweise lässt sich dazu das Kategoriensystem zur Auswertung der Stellenanzeigen nutzen.

Stellenwert eigenschaftstheoretischer Ansätze

Verbunden mit diesen methodischen Ansätzen wurde auch die inhaltliche Frage behandelt, welchen Beitrag Eigenschaftskonstrukte und die Eignungsdiagnostik für die Führungsforschung und die Managementdiagnostik leisten können. Solche Fragen wurden bisher im Rahmen von Validierungsstudien induktiv bearbeitet, weniger jedoch konzeptuell. Dennoch lässt sich aus der Befundlage anderer und der eigenen empirischen Studien schließen, dass es leistungs- und nicht leistungsbezogene Dispositionen gibt, die mit Führungserfolg zusammenhängen. Die Korrelationen liegen nach Cohen (1988) im mittleren Bereich, sind also nicht unerheblich und größer als zum Beispiel die Effektstärken in den meisten Trainingsevaluationen.

Gleichzeitig lassen diese mittleren Effekte reichlich Spielraum für den Einfluss weiterer Wirkprozesse, die vermutlich in anderen Führungstheorien, wie zum Beispiel situativen, interaktionalen oder systemischen Ansätzen abgebildet sind. Wollte man den zusätzlichen Einfluss solcher Mechanismen sowie die Wechselwirkung mit dispositionellen Faktoren erforschen, wäre nicht nur eine sorgfältige Diagnostik der Eigenschaftskonstrukte nötig, sondern ebenso der angesprochenen Prozesse.

Auch wenn viele Evaluationsstudien zur Trainingswirksamkeit enttäuschende Resultate ergaben, ist dennoch höchst plausibel, dass fachliche und nichtfachliche Weiterbildungen, Seminare, Kurse und so weiter Effekte haben, da Lernfähigkeit eine grundlegende menschliche Eigenschaft ist. Möglicherweise bildet eine differenziertere und präzisere Diagnostik der durch das Training angesprochenen Personenvariablen, die zudem Aggregations- und inhaltliche Symmetrien von Prädiktoren und Kriterien berücksichtigt, eine tragfähige Basis für weitere Evaluationsstudien (Wittmann, 1990). Dies erfordert nicht nur empirische Studien, sondern auch weitere konzeptuelle Arbeit am Inhalt und den Charakteristiken von Eigenschaften.

Zum Beispiel sollte auch geklärt werden, welche Persönlichkeitseigenschaften wie stark veränderlich sind. So wird meist davon ausgegangen, dass Intelligenz und Persönlichkeit im engeren Sinne sehr stabil seien, während etwa Motivation sich manipulieren lasse. Eine ambivalente Rolle spielt hier die sogenannte soziale Kompetenz oder interpersonelle Fertigkeiten. Einerseits soll sie ähnlich wie Intelligenz eine individuelle Begabung mit stabiler Aus-

prägung darstellen, andererseits zielen viele Trainings, insbesondere Team- und Führungskräftetrainings auf die Verbesserung derselben.

Fazit und Ausblick

Die eigenen Studien zur Managementdiagnostik und deren Evaluationsgesichtspunkten enthielten beispielhafte Anwendungen, hauptsächlich von Praxisstudien. Das Evaluationsmodell ist keine psychologische Theorie im engeren Sinne, da es sich nicht auf psychische Prozesse bezieht, weshalb sie weniger durch empirische Untersuchungen geprüft, sondern beispielhaft umgesetzt werden kann. Vielmehr handelt es sich um eine Klassifikation von Evaluationsstudien. Sicherlich wäre eine Veranschaulichung für die restlichen Bereiche auch wertvoll, doch da Evaluationsstudien von Praxisseite beauftragt werden, lässt sich dies von wissenschaftlicher Seite nicht erzwingen.

Welche der bekannten eignungsdiagnostischen Verfahren sich für die Managementdiagnostik besonders anbieten, ist unter Berücksichtigung der Zielgruppe und der Anwendungsintention zu entscheiden. So lassen sich aus den Ergebnissen der literaturgestützten Anforderungsanalyse Vorschläge für die inhaltliche Gestaltung der Verfahren ableiten, nicht aber für die Entscheidung für oder gegen eine bestimmte Methode, sofern sie generell den methodischen Anforderungen an diagnostische Verfahren genügt. Deshalb sind generell günstige methodische Eigenschaften der Verfahren und wichtige Konstrukte für die Managementdiagnostik zu ermitteln. Solche Befunde können in normativen Treatment-Evaluationen erhoben werden, und in eine normative Ziel-Evaluation der Managementdiagnostik einfließen.

Auch die Anwendung des Problemlösemodells auf die Inhalte von Führungsaufgaben sowie auf die berufliche Eignungsdiagnostik könnte dahin gehend präzisiert werden, dass sowohl konzeptuell wie empirisch die Überschneidungsbereiche des interpersonellen Bereichs verglichen werden. Interessant wären etwa verschiedene Varianten computergestützter Szenarios sowie ein Vergleich verschiedener Verfahren, etwa Assessment-Center-Übungen, die das Problemlösemodell implementieren. Dafür liefert die vorliegende Arbeit einen weiteren Zugang und vielleicht einen Anstoß zu einer differenzierteren Auseinandersetzung mit der Frage, inwieweit sich Führungstätigkeit als komplexes Problemlösen betrachten lässt.

Generell lässt sich daraus schließen, dass eine fundierte Kompetenzdiagnostik, gleich mit welchen Verfahren, nicht nur methodisch sorgfältig durchgeführt werden muss, sondern gleichzeitig auf eigenschaftstheoretischen Konstrukten beruhen sollte, um einen Beitrag zu einer substanziellen und begründeten Zielprognose von Führungserfolg zu liefern. Auch Führungserfolg sollte vor diesem Hintergrund modelliert werden.

Literaturverzeichnis

- Anderson, J. R. (1982). Aquisition of cognitive skill. *Psychological Review*, 89, 369-406.
- Antoni, C.H. (1993). Evaluationsforschung in der Arbeits- und Organisationspsychologie. In W. Bungard & T. Herrmann (Hrsg.), *Arbeits- und Organisationspsychologie im Spannungsfeld zwischen Grundlagenforschung und Anwendung*, 309-337. Bern: Huber.
- Arthur, W., Woehr, D.J. & Maldegen, R. (2000). Convergent and discriminant validity of assessment center dimensions: A conceptual and empirical reexamination of the assessment center construct related validity paradox. *Journal of Management*, 26(4), 813-835.
- Arvey, R.D. & Murphy, K.R. (1998). Performance evaluation in work settings. *Annual review of psychology*, 49, 141-168.
- Ashton, M.C. (1998). Personality and job performance: The importance of narrow traits. *Journal of Organizational Behavior*, 19, 289-303.
- Badke-Schaub, P. (2001). Erkennen und Bewältigen kritischer Situationen in Projektgruppen. In: R. Fisch, D. Beck, & B. Englich (Hrsg.), *Projektgruppen in Organisationen. Praktische Erfahrungen und Erträge der Forschung*, 249-267. Göttingen: Hogrefe.
- Bandura, A. (1986). *Social foundations of thought and action*. Englewood Cliffs: Prentice Hall.
- Barick, M.R. & Mount, M.K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26.
- Baron-Boldt, J., Funke, U. & Schuler, H. (1989). Prognostische Validität von Schulnoten. Eine Metaanalyse der Prognose des Studien- und Ausbildungserfolgs. In R.S. Jäger, R. Horn & K. Ingenkamp (Hrsg.), *Tests und Trends* 7, 11-39. Weinheim: Beltz.
- Barrett, G.V. (1992). Clarifying construct validity: Definitions, processes, and models. *Human Performance*, 5, 13-58.

- Bartels, G. (1987). Personalwerbung für Führungskräfte. In A. Kieser, G. Reber und R. Wunderer (Hrsg.). Handwörterbuch der Führung, 1695-1701. Stuttgart: Poeschel.
- Bass, B.M. (1985). Leadership and performance beyond expectations. New York: Free Press.
- Bass, B.M. (1990). Bass and Stogdill's handbook of leadership. Theory, research, and managerial applications. New York: Free Press.
- Bentler, P.M. (1998). EQS for Windows 5.7b. Multivariate Software, Inc.
- Beyer, H.T. (1970). Die Lehre der Unternehmensführung. Berlin.
- Biedermann, C. (1989). Subjektive Führungstheorien. Bern: Haupt.
- Blake, R.R. & Mouton, J.S. (1964). The managerial grid. Houston: Gulf Publishing.
- Blickle, G. & Hepperle, S. (1999). Strukturgleichungsmodelle zur Analyse der Effekte von Zielen, Motiven und Machtbasen bei der Beeinflussung von Vorgesetzten und KollegInnen. Zeitschrift für Arbeits- und Organisationspsychologie, 43, 26-35.
- Bobko, P., Roth, P.L. & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. Personnel Psychology, 52, 561-589.
- Borman, W.C. & Brush, D.H. (1993). More progress toward a taxonomy of managerial performance requirements. Human Performance, 6(1), 1-21.
- Borman, W.C. & Motowidlo, S.J. (1997). Task performance and contextual performance: The meaning for personnel selection research. Human Performance, 10, 99-109.
- Borman, W.C., Hanson, M.A. & Hedge, J.W. (1997). Personnel selection. Annual Review of Psychology, 299-337.
- Boudreau, J.W., Boswell, W.R. & Judge, T.A. (2001). Effects of personality on executive career success in the United States and Europe. Journal of Vocational Behavior, 58, 53-81.
- Brandtstädter, J. (1990). Evaluationsforschung: Probleme der wissenschaftlichen Bewertung von Interventions- und Reformprojekten. Zeitschrift für Pädagogische Psychologie, 4, 215-227.
- Briggs, K.C. & Briggs-Myers, I. (1991). Myers-Briggs Typenindikator. MB-TI. Testmappe. Deutsche Bearbeitung von Richard Bents und Reiner Blank. Weinheim: Beltz.

- Bröder, A. (2000). Kommentar zur Methodologie der deskriptiven Entscheidungsforschung. *Psychologische Beiträge*, 42, 645-662.
- Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Reviews*, 62, 193-217.
- Buchner, A. (1999). Komplexes Problemlösen vor dem Hintergrund finiter Automaten. *Psychologische Rundschau*, 50, 206-212.
- Bungard, W. (1987). Zur Problematik von Reaktivitätseffekten bei der Durchführung eines Assessment Centers. In H. Schuler & W. Stehle (Hrsg.), *Assessment Center als Methode der Personalentwicklung*, 99-125. Stuttgart: Verlag für Angewandte Psychologie.
- Burns, J.M. (1978). *Leadership*. New York: Harper & Row.
- Campbell, D.T. & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Carless, S.A. (1998). Assessing the discriminant validity of transformational leader behaviour as measured by the MLQ. *Journal of Occupational and Organizational Psychology*, 71, 353-358.
- Carless, S.A. & Allwood, V.E. (1997). Managerial assessment centres: What is being rated? *Australian Psychologist*, 32(2), 101-105.
- Carroll, S.J. & Gillen, D. (1987). Are the classical management functions useful in describing managerial work? *Academy of Management Review*, 12(1), 38-51.
- Casti, J. (1979). *Connectivity, complexity and catastrophe in large-scale systems*. New York: Wiley.
- Chen, H.T. (1990). *Theory driven evaluations*. Newbury Park: Sage.
- Chi, M.T.H., Glaser, R. & Farr, M.J. (1988). *The nature of expertise*. Hillsdale: Lawrence Erlbaum Associates.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale: Erlbaum.
- Cohen, J. & Cohen, P. (1983). *Applied multiple regression: Correlational analysis for behavioral sciences*. Hillsdale: Erlbaum.
- Conger, J.A. & Kanungo, R. (1987). Toward a behavioral theory of charismatic leadership in organizational settings. *Academy of Management Review*, 12, 637-647.

- Connelly, M.S., Gilbert, J.A., Zaccaro, S.J., Threlfall, K.V., Marks, M.A. & Mumford, M.D. (2000). Exploring the relationship of leadership skills and knowledge to leader performance. *Leadership Quarterly*, 11(1), 65-86.
- Conrad, W. & Michel, L. (1982). Theoretische Grundlagen psychometrischer Tests. In K.J.M. Groffmann (Hrsg), *Enzyklopädie der Psychologie*, B, II, 1, 1-129. Göttingen: Hogrefe.
- Conrad, W., Mohr, V. & Seydel, U. (1980). Analyse und Skalierung der 16-PF-Form E (von Cattell). *Zeitschrift für Differentielle und Diagnostische Psychologie*, 1(2), 169-182.
- Conway, J.M. (1999). Distinguishing contextual performance from task performance for managerial jobs. *Journal of Applied Psychology*, 84(1), 3-13.
- Conway, J.M. & Huffcutt, A.I. (1999). Psychometric properties of multi-source performance ratings: A meta-analysis of subordinate, supervisor, peer and self-ratings. *Human Performance*, 10(4), 331-360.
- Cook, T. (1998). *Personnel selection: Adding value through people*. Chichester: Wiley.
- Cook, T.D. & Wittmann, W.W. (1998). Lessons learned about evaluation in the United States and some possible implications for Europe. *European Journal of Psychological Assessment*, 14, 97-115.
- Cook, T.D. Campbell, D.T. (1979). *Quasi-Experimentation. Design and analysis issues for field settings*. Chicago: Rand McNally.
- Cooper, H. & Hedges, L.V. (1984). *The handbook of research synthesis*. New York: Sage.
- Cortina, J.M., Goldstein, N.B., Payne, S.C., Davison, H.K. & Gilliland, S.W. (2000). The incremental validity of interview scores over and above cognitive ability and conscientiousness scores. *Personnel Psychology*, 53, 325-351.
- Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.
- De Corte, W. (1999). Weighing job performance predictors to both maximize the quality of the selected workforce and control the level of adverse impact. *Journal of Applied Psychology*, 84(5), 695-702.
- Der Stellenreport. Wintersemester 2000/2001. Tegernsee: Verlagsgesellschaft mbH.

- Doll, E.A. (1953) The measurement of social competence: A manual for the Vineland Social Maturity Scale. Minnesota: Educational Test Bureau.
- Donahue, L.M., Truxillo, D.M., Conrwell, J.M. & Gerrity, M.J. (1997). Assessment center construct validity and behavioral checklists: Some additional findings. *Journal of Social Behavior and Personality*, 12(5), 85-108.
- Dörner, D. (1976). Problemlösen als Informationsverarbeitung. Stuttgart: Kohlhammer.
- Dörner, D. (1986). Diagnostik der operativen Intelligenz. *Diagnostica*, 32, 290-308.
- Evans, M.G. (1970). The effects of supervisory behavior on path-goal relationship. *Organizational Behavior and Human Performance*, 5, 277-298.
- Fayol, H. (1929). Allgemeine und industrielle Verwaltung. Berlin.
- Fiedler, F.E. (1967). A theory of leadership effectiveness. New York: McGraw-Hill.
- Fiedler, F.E. & Garcia, J.E. (1987). New approaches to leadership: Cognitive resources and organizational performance. New York: Wiley.
- Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327-358.
- Fleenor, J.W. (1996). Constructs and developmental assessment centers: Further troubling empirical findings. *Journal of Business and Psychology*, 10(3), 319-335.
- Fleishman, E.A., Mumford, M.D., Zaccaro, S.J., Levin, K.Y., Korotkin, A.L. & Hein, M.B. (1991). Taxonomic efforts in the description of leadership behavior: A synthesis and a functional interpretation. *Leadership Quarterly*, 2(4), 245-287.
- French & Raven (1959). The bases of social power. In D. Cartwright (Hrsg.), *Studies of social power*, 150-167. Ann Arbor: Institute for social research.
- Fricke, R. & Treinies, G. (1985). Einführung in die Metaanalyse. Bern: Huber.
- Frieling, E. (1975). Psychologische Arbeitsanalyse. Stuttgart: Kohlhammer.
- Früh, W. (1998). Inhaltsanalyse. Theorie und Praxis. Konstanz: UVK Medien.

- Funke, U. (1992). Die Validität einer computergestützten Systemsimulation zur Diagnose von Problemlösekompetenz. In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis*. Stuttgart: Verlag für Angewandte Psychologie.
- Funke, U. (i.V.). *Arbeitsanalyse und Eignungsdiagnostik: Konzepte und Anwendungen am Beispiel industrieller Forschung und Entwicklung*. Universität Hohenheim: Dissertation.
- Funke, U. & Schuler, H. (1998). Validity of stimulus and response components in a video test of social competence. *International Journal of Selection and Assessment*, 1998, 6, 115-123
- Gebert, D. (1995). Ist eine Vertrauensorganisation überhaupt möglich? *Management Zeitschrift*, 10, 66-70.
- Geilhardt, T. & Mühlbradt, T. (1995). *Planspiele im Personal- und Organisationsmanagement*. Göttingen: Verlag für Angewandte Psychologie.
- Ghiselli, E.E. (1973). The validity of aptitude tests in personnel selection. *Personnel Psychology*, 26, 461-477.
- Giese, F. (1924). Die Arbeitsprobe in der Psychodiagnostik. *Zeitschrift für Angewandte Psychologie*, 23, 3-4.
- Goldman, S.R., Zech, L.K., Biswas, G. & Noser, T. (1999). Computer technology and complex problem solving: Issues in the study of complex cognitive activity. *Instructional Science*, 27, 235-268.
- Goodstein, L.D. & Lanyon, R.I. (1999). Applications of personality assessment to the workplace: A review. *Journal of Business Psychology*, 13 (3), 291-322.
- Gulick, L.H. (1937). Notes on the theory of organization. In L.H. Gulick & Urwick, L.F. (Hrsg.), *Papers on the science of administration*. New York.
- Guthke, J. (1982). The learning test concept - An alternative to the traditional static intelligence test. *The German Journal of Psychology*, 6, 306-324.
- Hacker, W. (1983). Psychische Anforderungen. In F. Stoll (Hrsg.), *Arbeit und Beruf. Kindlers Psychologie des 20. Jahrhunderts*, 86-109. Weinheim: Beltz.
- Haeberlin, F. (1995). Interessen. In W. Sarges (Hrsg.), *Management-Diagnostik*, 346-353. Göttingen: Hogrefe.

- Hank, P. & Schwenkmezger P. (1992). Myers-Briggs Typenindikator (MB-TI) nach C.C. Briggs und I. Briggs Myers (Testrezension). Zeitschrift für Differentielle und Diagnostische Psychologie, 13, 130-132.
- Hasselmann, D. (1993). Computersimulierte komplexe Problemstellungen der Management-Diagnostik. Hamburg: Windmühle.
- Heckhausen, H., Schmalt, H.D. & Schneider, K. (1985). Achievement motivation in perspective. New York: Academic Press.
- Hemphill, J.K. (1960). Dimensions of executives positions. Research Bulletin RB-59-5. Princeton: ETS.
- Hermelin, E. & Robertson, I.T. (2001). A critique and standardization of meta-analytic validity coefficients in personnel selection. Journal of Occupational and Organizational Psychology, 74, 253-277.
- Herrmann, T. (1991). Lehrbuch der empirischen Persönlichkeitsforschung. Göttingen: Hogrefe.
- Hersey, P. & Blanchard, K.H. (1969). Life cycle theory of leadership. Training and Development Journal, 23(2), 4-26.
- Hoffmann, C.C. & Thornton, G.C. (1997). Examining selection utility where competing predictors differ in adverse impact. Personnel Psychology, 50, 455-470.
- Hofstätter, P.R. (1986). Gruppendynamik: Kritik der Massenpsychologie. Reinbek: Rowohlt.
- Höft, S. (2001). Erfolgsüberprüfung personalpsychologischer Arbeit. In: H. Schuler (Hrsg.), Lehrbuch Personalpsychologie, 618-651. Göttingen: Hogrefe.
- Hogan, R., Curphy, G.J., & Hogan, J. (1994). What we know about leadership: Effectiveness and personality. American Psychologist, 49, 493-504.
- Holling, H. & Gediga, G. (1999). Evaluationsforschung. Göttingen: Hogrefe.
- Hough, L.M. (1992). The big five personality variables - construct confusion: description vs. prediction. Human Performance, 5, 139-155.
- Hough, L.M. & Oswald, F.L. (2000). Personnel selection: Looking toward the future - remembering the past. Annual Review of Psychology, 51, 631-664.
- House, R.J. & Howell, J.M. (1992). Personality and charismatic leadership. Leadership Quarterly, 3, 81-108.

- Huffcutt, A.I., Roth, P.L. & McDaniel, M.A. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology*, 81(5), 459-473.
- Hunt, J.G. (1991). *Leadership: A new synthesis*. Newbury Park: Sage.
- Hunter, J.E. & Hunter, R.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96(1), 72-98.
- Hunter, J.E. & Schmidt, F.L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks: Sage Publications.
- Jäger, R.S. (1982). Diagnostische Urteilsbildung. In K.J.M. Groffmann (Hrsg.), *Enzyklopädie der Psychologie*, B, II, 1, 295-375. Göttingen: Hogrefe.
- Jäger, R.S. (1992). Biografische Daten. In R.S. Jäger & F. Petermann (Hrsg.), *Psychologische Diagnostik*, 158-161. Weinheim: Psychologie Verlags-Union.
- Jeserich, W. (1981). *Mitarbeiter auswählen und fördern*. München: Hanser.
- Jochmann, W. (1992). Von der Personalentwicklung zur Unternehmensentwicklung. *Personal*, 44 (9), 410-413.
- Jones, A., Herriot, P., Long, B. & Drakely, R. (1991). Attempting to improve the validity of a well-established assessment centre. *Journal of Occupational Psychology*, 64, 1-21.
- Joyce, L.W., Thayer, P.W. & Pond, S.B. (1994). Managerial functions: An alternative to traditional assessment center dimensions? *Personnel Psychology*, 47, 109-121.
- Jungermann, H. (1995). Entscheiden. In W. Sarges (Hrsg.), *Management-Diagnostik*, 141-150. Göttingen: Hogrefe.
- Kahn, R.L., Wolfe, D.M., Quinn, R.P. & Snoek, J.D. (1964). *Organizational stress: Studies in role conflict and ambiguity*. New York: Wiley.
- Kannheiser, W. (1995). Erfassung der Anforderungen einer konkreten Position. In W. Sarges (Hrsg.), *Management-Diagnostik*, 141-150. Göttingen: Hogrefe.
- Kaplan, R.E. (1986). The warp and woof of the general manager's job. Technical Report No. 27, 1-32. Greensboro: Center for Creative Leadership.
- Karkoschka, U. (1997). *Validität eignungsdiagnostischer Verfahren zur Messung sozialer Kompetenz*. Frankfurt: Lang.

- Katz, R.L. (1955). Skills of an effective administrator. *Harvard Business Review*, 33-42.
- Kerr, S. & Jermier, J.M. (1978). Substitutes for leadership: Their meaning and measurement. *Organizational Behavior and Human Performance*, 22, 375-403.
- Kersting, M. (1999). Diagnostik und Personalauswahl mit computergestützten Problemlöseszenarien? Göttingen: Hogrefe.
- Kessler, B.H. (1988). Daten aus dem Interview. In R.S. Jäger (Hrsg.), *Psychologische Diagnostik - Ein Lehrbuch*, 363-372. München: Psychologie Verlags Union.
- Kickul, J. & Neuman, G. (2000). Emergent leadership behaviors: The function of personality and cognitive ability in determining teamwork performance and ksas. *Journal of business and psychology*, 15, 2-51.
- Kieser, A. (1997). Stichwort Organizational Evolution. In: *The Handbook of Organizational Behaviour* A. Sorge & M. Warner, 580-587. London.
- Kirkpatrick, D.L. (1987). Evaluation of training. In R.L. Craig (Hrsg.), *Training and development handbook: A guide to human resource development*, 301-319. New York: McGraw-Hill.
- Kirkpatrick, S.A. & Locke, E.A. (1991). Leadership: Do traits matter? *The Academy of Management Executive*, 5(2), 48-60.
- Kleinmann, M. (1997). *Assessment-Center*. Göttingen: Verlag für Angewandte Psychologie.
- Klenke, K. (1994). Meta-analytic studies of leadership: Added insights or added paradoxes? *Current Psychology*, 12(4), 326-343.
- Kluwe, R.H. (1995). Computergestützte Systemsimulationen. In W. Sarges (Hrsg.), *Management-Diagnostik*, 218-225. Göttingen: Hogrefe.
- Kluwe, R.H. (1995a). Wissen. In W. Sarges (Hrsg.), *Management-Diagnostik*, 218-225. Göttingen: Hogrefe.
- Kluwe, R.H., Schilde, A., Fischer, C. & Oellerer, N. (1991). Problemlöseleistungen beim Umgang mit komplexen Systemen und Intelligenz. *Diagnostica*, 37, 291-313.
- Knoblauch, R. (1990). Die Auswahl von Außendienstmitarbeitern in der pharmazeutischen Industrie. In H. Schuler und W. Stehle (Hrsg.), *Biografische Fragebogen als Methode der Personalauswahl*, 91-113. Göttingen: Verlag für Angewandte Psychologie.
- Kotter, J. (1991) *Abschied vom Erbsenzähler (Leadership: A force for change)*. Düsseldorf: Econ.

- Küpper, W. & Felsch, A. (2000). Organisation, Macht und Ökonomie: Mikropolitik und die Konstitution organisationaler Handlungssysteme. Opladen: Westdeutscher Verlag.
- Küpper, W. & Ortmann, G. (1988): Mikropolitik. Rationalität, Macht und Spiele in Organisationen. Opladen: Westdeutscher Verlag.
- Larson, G.E. & Wolfe, J.H. (1995). Validity results for g from an expanded test. *Intelligence*, 20, 15-25.
- Lasser, R. (1987). Symbolische Führung. In A. Kieser, G. Reberer & R. Wunderer (Hrsg.), *Handwörterbuch der Führung, 1927-1938*. Stuttgart: Poeschel.
- Latham, G.P., Saari, L.M., Pursell, E.D. & Campion, M.A. (1980). The situational interview. *Journal of Applied Psychology*, 54, 417-423.
- Lattmann, C. (1982). Die verhaltenswissenschaftlichen Grundlagen der Führung der Mitarbeiter. Bern: Huber.
- Laukamm, T. & Walsh, L. (1985). Strategisches Management von Human-Ressourcen. In Arthur D. Little International (Hrsg.), *Management im Zeitalter der strategischen Führung*, 80-100. Wiesbaden: Gabler.
- Lewin, K., Lippitt, R. & White, R.K. (1939). Patterns of aggressive behavior in experimentally created social climates. *Journal of Social Psychology*, 10, 271-301.
- Luhman, N. (1988). Organisation. In W. Küpper & G. Ortmann (Hrsg.), *Mikropolitik: Rationalität, Macht und Spiele in Organisationen*, 165-185. Opladen: Westdeutscher Verlag.
- Luhmann, N. (1984). *Soziale Systeme*. Frankfurt: Suhrkamp.
- Luthans, F. & Lockwood, D.L. (1984). Toward an observation system for measuring leader behavior in natural settings. In J.G. Hunt, D. Hosking, C.A. Schriesheim & R. Stewart (Hrsg.), *Leaders and managers: International perspectives on managerial behavior and leadership*. New York: Pergamon Press.
- Malik, F. (1986). *Strategien des Managements komplexer Probleme*. Bern: Haupt.
- Marsden, P.V. (1994). Selection methods in US establishments. *Acta Sociologica*, 37, 287-301.
- Marshall-Mies, J.C., Fleishman, E.A., Martin, J.A., Zaccaro, S.J., Baughman W.A. & McGee, M.L. (2000). Development and evaluation of cognitive and metacognitive measures for predicting leadership potential. *Leadership Quarterly*, 11(1), 135-153.

- Matarazzo, J.D. (1972). Wechsler's measurement and appraisal of adult intelligence. Baltimore, Williams and Wilkins.
- Mayring, P. (1997). Qualitative Inhaltsanalyse: Grundlagen und Techniken. Weinheim: Deutscher Studien-Verlag.
- McClelland, D.C. & Boyatzis, R.E. (1982). Leadership motive pattern and long term success in management. *Journal of Applied Psychology*, 67, 737-743.
- McCrae, R.R. & Costa, P.T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52, 81-90.
- McFall, R.M. (1982). A review and reformulation of the concept of social skills. *Behavioral Assessment*, 4, 1-33.
- Meehl, P.E. (1954). Clinical vs. statistical prediction. Minneapolis: University of Minnesota Press.
- Meffert, H. & Wagner, H. (1992). Qualifikation und Ausbildung von Führungskräften: Empirische Befunde und Implikationen. *Zeitschrift für Personalforschung*, 3, 352-265.
- Miner, J.B. (1990). Entrepreneurs, high growth entrepreneurs, and managers: Contrasting and overlapping motivational patterns. *Journal of Business Venturing*, 5, 221-234.
- Mintzberg, H. (1973). The nature of managerial work. New York: Harper and Row.
- Motowidlo, S.J., Burnett, J.R., Maczynski, J., Witkowski, S., Wojtachnio, A. & Chelupa, S. (1996). Predicting managerial job performance from personality ratings based on a structured interview: An international replication. *Polish Psychological Bulletin*, 27(2), 139-151.
- Müller, G.F. (1999). Persönlichkeit und selbständige Erwerbstätigkeit: Messung von Potentialfaktoren und Analyse ihres Einflusses. *Wirtschaftspsychologie*, 1999, 6 (2), 2-12.
- Mumford, M.D., Zaccaro, S.J., Harding, F.D., Jacobs, T.O. & Fleishman, E.A. (2000). Leadership skills for a changing world: Solving complex social problems. *Leadership Quarterly*, 11, 11-35.
- Murphy, K.H. & Shiarella, A.H. (1997). Implications of the multidimensional nature of job performance for the validity of personnel selection tests: Multivariate framework for studying test validity. *Personnel Psychology*, 50, 823-854.

- Musch, J. & Lieberei, W. (1997). Entwicklung und Validierung einer auswertungsobjektiven Postkorbübung für die Personalauswahl. *Zeitschrift für Personalforschung*, 11, 317-336.
- Nachreiner, F. & Müller, G.F. (1995). Führungsmotivation. In W. Sarges (Hrsg.), *Management-Diagnostik*, 271-277. Göttingen: Hogrefe.
- Neuberger, O. (1995). Mikropolitik. Der alltägliche Aufbau und Einsatz von Macht in Organisationen. Stuttgart: Lucius und Lucius.
- Neuberger, O. (2002). Führen und Führen lassen. Stuttgart: Lucius und Lucius.
- Neuberger, O. (1990). Führen und geführt werden. Stuttgart: Enke.
- Neuberger, O. (2000). Das 360-Grad Feedback. München: Hampp.
- Neuberger, O. (2000a). Vom Helden zum Opfer: Organisationspsychologe Oswald Neuberger über Topmanager, Fusionen und Koalitionen. *DIE ZEIT*, 16.
- Niederfeichtner, F. (1987). Auswahl von Führungskräften. In A. Kieser, G. Reber & R. Wunderer (Hrsg.), *Handwörterbuch der Führung*, 82-90. Stuttgart: Poeschel.
- Ones, D.S. & Viswesvaran, C. (1996). Bandwidth-fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior*, 17, 609-626.
- Ones, D.S., Mount, M.K., Barrick, M.R. & Hunter, J.E. (1994). Personality and job performance: A critique of the Tett, Jackson, and Rothstein (1991) meta-analysis. *Personnel Psychology*, 47, 147-156.
- Orth, B. (1974). Einführung in die Theorie des Messens. Stuttgart: Kohlhammer.
- Paunonen, S.V., Rothstein, M.G. & Jackson, D.N. (1999). Narrow reasoning about the use of broad personality measures for personnel selection. *Journal of Organizational Behavior*, 20, 389-405.
- Pavett, C.M. & Lau, A.W. (1983). Managerial work: The influence of hierarchical level and functional specificity. *Academy of Management Journal*, 26, 170-177.
- Pawlik, K. (1976). Diagnose der Diagnostik. Stuttgart: Klett.
- Popper, K. (1976). Logik der Forschung. Tübingen: Mohr Siebeck.
- Putz-Osterloh (1995). Problemlösen. In W. Sarges (Hrsg.), *Management-Diagnostik*, 240-247. Göttingen: Hogrefe.

- Ramme, I. (1990). Die Arbeit von Führungskräften: Konzepte und empirische Ergebnisse. Bergisch-Gladbach: Josef Eul.
- Ree, M.J., Earles, J.A. & Teachout, M.S. (1994). Predicting job performance: Not much more than g. *Journal of Applied Psychology*, 80, 721-730.
- Reilly, R.R. & Chao, G.T. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology*, 35, 1-62.
- Robertson, I.T. & Kandola, R.S. (1983). Work sample tests: Validity, adverse impact and applicant reaction. *Journal-of-Occupational-Psychology*, 55(3), 171-183.
- Robertson, I.T. & Makin, P.J. (1986). Management selection in Britain. A survey. *Journal of Occupational Psychology*, 59, 45-57.
- Robertson, I.T. & Smith, M. (2001). Personnel selection. *Journal of Occupational and Organizational Psychology*, 74, 441-472.
- Robertson, I.T., Baron, H., Gibbons, P., MacIver, R. & Nyfield, G. (2000). Conscientiousness and managerial performance. *Journal of occupational and organizational psychology*, 73, 171-180.
- Rosenstiel, L. (1992). Grundlagen der Organisationspsychologie. Stuttgart: Poeschel.
- Rosenstiel, L. (1994). Führung im Systemwandel. Untersuchungen zum Führungsverhalten beim Übergang von der Plan- in die Marktwirtschaft. München: Hampp.
- Rosenstiel, L. (1995). Werthaltungen. In W. Sarges (Hrsg.), *Management-Diagnostik*, 329-335. Göttingen: Hogrefe.
- Rosenstiel, L. (2001). Führung. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie*, 317-347.
- Rosenthal, R. & DiMatteo, M.R. (2001). Meta-Analysis: Recent Development in quantitative methods for literature reviews. *Annual review of psychology*, 52, 59-82.
- Rossi, P.H, Freeman, H.E. & Lipsey, M.W. (1999). *Evaluation A Systematic Approach*. New York: Sage
- Rossi, P.H. & Freeman, H.E. (1993). *Evaluation: A systematic approach*. London: Sage.
- Runde, B., Kirschbaum, D. & Wübbelmann, K. (2001). 360-Feedback - Hinweise für ein best-practice-Modell. *Zeitschrift für Arbeits- und Organisationspsychologie*, 45, 146-157.

- Russell, C.G. (1987). Person characteristics versus role congruency explanations for assessment centre ratings. *Academy of Management Journal*, 30, 817-826.
- Sackett, P.R. & Dreher, G.F. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. *Journal of Applied Psychology*, 67, 401-110.
- Sackett, P.R., Gruys, M.L. & Ellingson, J.E. (1998). Ability-personality interactions when predicting job performance. *Journal of Applied Psychology*, 83 (4), 545-556.
- Sarges, W. (1995). Lernpotential-AC. In W. Sarges (Hrsg.), *Management-Diagnostik*, 728-739. Göttingen: Hogrefe.
- Sarges, W. (1995). *Management-Diagnostik*. Göttingen: Hogrefe.
- Schieffer, A. (1998). *Führungspersönlichkeit: Struktur, Wirkung und Entwicklung erfolgreicher Top-Führungskräfte*. Wiesbaden: Gabler.
- Schilling, J. (2001). *Wovon sprechen Führungskräfte, wenn sie über Führung sprechen? Eine Analyse subjektiver Führungstheorien*. Hamburg: Dr. Kovac.
- Schirmer, F. (1992). *Arbeitsverhalten von Managern. Bestandsaufnahme, Kritik und Weiterentwicklung der Aktivitätsforschung*. Wiesbaden: Gabler.
- Schmidt, F. & Hunter, J. (1998). The validity and utility of selection methods in personnel psychology. Practical and theoretical implications of 85 years of research findings. *Psychological bulletin*, 124, 262-274.
- Schmidt, F.L. & Hunter, J.E. (1977). Development of a general solution to the problem of validity generalisation. *Journal of Applied Psychology*, 62, 529-540.
- Schmidt, H. (1981). *MPT-E: Mehrdimensionaler Persönlichkeitstest für Erwachsene*. Braunschweig: Westermann.
- Schmidt-Klingenberg, M. (2002). Unter Druck nach oben. *Bildungsreihe Teil 9: Karriere. DER SPIEGEL*, 28, 58-66.
- Schmitt, N., Gooding, R.Z., Noe R.D. & Kirsch, M. (1984). Metaanalysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, 37, 407-422.
- Schneider, Ackerman und Kanfer (1997). To „act wisely in human relations.“ Exploring the dimensions of social competence. *Personal and individual Differences*, 21 (4), 469-481.

- Schuler, H. (1989). Some advantages and problems of job analysis. In M. Smith und I. Robertson (Hrsg.), *Advances in selection and assessment*, 31-42. Chichester: Wiley.
- Schuler, H. (1990). Personalauswahl aus der Sicht der Bewerber: Zum Erleben eignungsdiagnostischer Situationen. *Zeitschrift für Arbeits- und Organisationspsychologie*, 43, 184-191.
- Schuler, H. (1992). Das Multimodale Einstellungsinterview. *Diagnostica*, 38 (4), 281-300.
- Schuler, H. (1996). *Psychologische Personalauswahl*. Göttingen: Verlag für Angewandte Psychologie.
- Schuler, H. & Funke, U. (1989). Berufseignungsdiagnostik. In E. Roth (Hrsg.). *Organisationspsychologie, Enzyklopädie der Psychologie*, D, III, 3, 281-320. Göttingen: Hogrefe.
- Schuler, H. & Funke, U. (1993). Diagnose beruflicher Eignung und Leistung. In H. Schuler, H. Brandstätter, W. Bungard, S. Greif, E. Ulich & B. Wilpert (Hrsg.), *Lehrbuch Organisationspsychologie*, 235-283, Bern: Huber.
- Schuler, H. & Prochaska, M. (2000). Entwicklung und Konstruktvalidierung eines berufsbezogenen Leistungsmotivationstests. *Diagnostica*, 46, 61-72
- Schuler, H., Frier, D. & Kaufmann, M. (1993). Personalauswahl im europäischen Vergleich. Göttingen: Hogrefe.
- Schulz, C., Schuler, H. & Stehle, W. (1985). Organisationspsychologie und Unternehmenspraxis: Perspektiven der Kooperation. In H. Schuler und W. Stehle (Hrsg.), *Beiträge zur Organisationspsychologie*, 1, 126-132. Hohenheimer Gesprächstage zur Organisationspsychologie. Stuttgart: Verlag für angewandte Psychologie.
- Scriven, M. (1991). *Evaluation thesaurus*. Thousand Oaks: Sage.
- Scriven, M. (1999). The fine line between evaluation and explanation. *Research on social work practice*, 9, 521-524.
- Sechrest, L. & Figueredo, A.J. (1993). Program evaluation. *Annual review of Psychology*, 44, 645-674.
- Shadish, W.R. (1986). Planned critical multiplism: Some elaborations. *Behavioral Assessment*, 8, 75-103.
- Staehle, W.H. (1999). *Management. Eine verhaltenswissenschaftliche Perspektive*. München: Vahlen.

- Starbuck, W.H. (1983). Organizations as action generators. *American Sociological Review*, 48, 91-103.
- Steinhoff, V. (1995) Vorgesetztenbeurteilung: Grundlagen - Philosophie - Anwendung. In: K. Hofmann, F. Köhler, & V. Steinhoff (Hrsg.). *Vorgesetztenbeurteilung in der Praxis. Konzepte, Analysen, Erfahrungen*, 7-14. Weinheim: PVU.
- Sternberg, R.J. & J.C. Kaufman (1998). Human Abilities. *Annual Review of Psychology*, 49, 479-502.
- Stevens, M.J. & Campion, M.A. (1994). The knowledge, skill and ability requirements for teamwork: Implications for human resource management. *Journal of Management*, 20, 503-530.
- Stewart, R. (1976). *Contrasts in management*. London: McGraw-Hill.
- Stogdill, R.M. (1963). *Manual for the leader behavior description questionnaire, Form XII*. Columbus: Ohio State University.
- Stogdill, R.M. (1974). *Handbook of leadership: A survey of the literature*. New York: Free Press.
- Stogdill, R.M. & Shartle, C.L. (1955). *Methods in the study of administrative leadership*. Columbus: Bureau of Business Research.
- Strauß, B. & Kleinmann, M. (1995). *Computersimulierte Szenarien in der Personalarbeit*. Göttingen: Verlag für Angewandte Psychologie.
- Stufflebeam, D.L. (2000). The CIPP model for evaluation. In D.L. Stufflebeam, G.F. Madaus, & T. Kellaghan, (Eds.), *Evaluation Models* (2nd ed.). Boston: Kluwer Academic Publishers.
- Süß, H.-M. (1996). *Intelligenz, Wissen und Problemlösen*. Göttingen: Hogrefe.
- Tett, R.P., Guterman, H.A., Bleier, A. & Murphy, P.J. (2000). Development and content validation of a „hyperdimensional“ taxonomy of managerial competence. *Human Performance*, 13 (3), 205-251.
- Tett, R.P., Jackson, D.N., Rothstein, M. & Reddon, J.R. (1994). Meta-analysis of personality-job performance relations: A reply to Ones, Mount, Barrick and Hunter (1994). *Personnel Psychology*, 47, 157-168.
- Thayer, P.W. (1992). Construct validity: Do we understand our criteria? *Human Performance*, 5, 97-108.
- Thorndike, E.L. (1920). Intelligence and its uses. *Harper's Magazine*, 140, 217-235.

- Thornton, G.C., Gaugler, B.B., Rosenthal, D.B. & Bentson, C. (1992). Die prädiktive Validität des AC - eine Metaanalyse. In H. Schuler & W. Stehle (Hrsg.), *Assessment-Center als Methode der Personalentwicklung*, 36-60. Stuttgart: Verlag für angewandte Psychologie.
- Türk, K. (1981). *Personalführung und soziale Kontrolle*. Stuttgart: Enke.
- Ulich, E. (1998). *Arbeitspsychologie*. Stuttgart: Poeschel.
- Van Fleet, D.D. (1975). Changing patterns of significant authors on leadership and managerial effectiveness. *Journal of Management*, 1, 39-44.
- Viswesvaran, C. & Ones, D.S. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology*, 48, 865-885.
- Vroom, V.H. & Yetton, P. (1973). *Leadership and decision making*. Pittsburgh: University of Pittsburgh Press.
- Wagener, D. (1994). *Entwicklung des computergestützten Systems FSYS zur Diagnose komplexen Problemlösens*. Mannheim: Diplomarbeit.
- Wagener, D. (2001). *Psychologische Diagnostik mit komplexen Szenarios: Taxonomie, Entwicklung, Evaluation*. Lengerich: Pabst.
- Wagner, R. K. & Sternberg, R. J. (1985). Practical intelligence in real world pursuits: The role of tacit knowledge. *Journal of Personality and Social Psychology*, 49, 436-458.
- Wainer, H.A. & Rubin, I.M. (1971). Motivation of research and development entrepreneurs: Determinants of company success. In D.A. Kolb, I.M. Rubin & I. McIntire (Hrsg.), *Organizational Psychology*, 131-139. New Jersey: Prentice Hall.
- Walker, R.E. & Foley, J.M. (1973). Social intelligence: Its history and measurement. *Psychological Reports*, 33, 839-864.
- Wallach, M.A. (1976). Tests tell us little about talent. *American Scientist*, 64, 57-63.
- Weber, M. (1956). *Wirtschaft und Gesellschaft*. Tübingen: Mohr.
- Wehrich, H. & Koontz, H. (1993). *Management: A Global Perspective*. Hightstown: McGraw-Hill.
- Weinstein, E.A. (1969). The development of interpersonal competence. In D.A. Goslin (Hrsg.), *Handbook of socialization theory and research*. Chicago: Rand McNally.
- Weiss, C. H. (2001). *Evaluation*. Upper Saddle River: Prentice Hall.

- Wiesner, W.H. & Cronshaw, S.F. (1988). A meta-analytic investigation of the impact of interview format and degree of structure on the validity of employment interview. *Journal of Occupational Psychology*, 61, 275-290.
- Wittmann, W.W. (1985). *Evaluationsforschung. Aufgaben, Probleme und Anwendungen*. Heidelberg: Springer.
- Wittmann, W.W. (1990). Brunswik-Symmetrie und die Konzeption der fünf Datenboxen - Ein Rahmenkonzept für umfassende Evaluationsforschung. *Zeitschrift für pädagogische Psychologie*, 4, 241-251.
- Wittmann, W.W. & Matt, G.E. (1986). Aggregation und Symmetrie. Grundlagen einer multivariaten Reliabilitäts- und Validitätstheorie, dargestellt am Beispiel der differentiellen Validität des BIS. *Diagnostica*, 32, 309-329.
- Wofford, J.C. (1982). An integrative theory of leadership. *Journal of Management*, 8, 27-47.
- Wofford, J.C. (1999). Laboratory Research on charismatic leadership: Fruitful or futile? *Leadership Quarterly*, 10(4), 523-529.
- Wortman, P.M. (1984). Judging research quality. In H. Cooper & L.V. Hedges (Hrsg.), *The handbook of research synthesis*, 97-110. New York: Sage.
- Wottawa, H. & Thierau, H. (1998). *Lehrbuch Evaluation*. Bern: Huber.
- Wren, D.A. (1990). Was Henry Fayol a real manager? *Academy of Management Proceedings*, 138-142.
- Wunderer, R. (1997). Führung des Chefs: Führung von unten durch sachliche Begründung und Freundlichkeit. *Personalführung*, 30, 148-157.
- Yukl, G. (1989). Managerial leadership: A review of theory and research. *Journal of Management*, 15, 251-289.
- Yukl, G. (1994). *Leadership in Organizations*. New Jersey: Prentice Hall.
- Yukl, G. (1999). An Evaluation of conceptual weaknesses in transformational and charismatic leadership theories. *Leadership Quarterly*, 10(2), 285-305.
- Yukl, G. (2002). *Leadership in Organizations*. New Jersey: Prentice Hall.
- Yukl, G. & Falbe, C.M. (1991). The importance of different power sources in downward and lateral relations. *Journal of Applied Psychology*, 76, 416-423.

- Yukl, G. & Van Fleet, D.D. (1992). Theory and Research on Leadership in Organizations. In M.D. Dunnette & L.M. Hough (Hrsg.), *Handbook of Industrial and Organizational Psychology*, 3, 147-197. Palo Alto: Consulting Psychologists Press.
- Zaccaro, S.J., Mumford, M.D., Connelly, M.S., Marks, M.A. & Gilbert, J.A. (2000). Assessment of leader problem solving capabilities. *Leadership Quarterly*, 11, 37-64.
- Zaleznik, A. (1977). Managers and leaders: Are they different? *Harvard Business Review*, 55, 67-78.
- Zimmer, D. (1978). Die Entwicklung des Begriffes der Selbstsicherheit und der sozialen Kompetenz in der Verhaltenstherapie. In R. Ullrich & R. Ullrich (Hrsg.), 2, *Soziale Kompetenz*. München: Pfeiffer. Projektgruppen in Organisationen. Praktische Erfahrungen und Erträge der Forschung, 249-267. Göttingen: Hogrefe.

Metaanalyse

- Boudreau, J.W., Boswell, W.R. & Judge, T.A. (2001). Effects of personality on executive career success in the United States and Europe. *Journal of Vocational Behavior*, 58, 53-81.
- Caligiuri, P.M. (2000). The big five personality characteristics as predictors of expatriate's desire to terminate the assignment and supervisor-rated performance. *Personnel psychology*, 53, 67-89.
- Church, A.H. & Waclawski, J. (1998). The relationship between individual personality orientation and executive leadership behaviour. *Journal of occupational and organizational psychology*, 71, 99-125.
- Conway, J.M. (2000). Managerial performance development: Constructs and personality correlates. *Human performance*, 13, 23-46.
- de Jong, R.D., Bouhuys, S.A. & Barnhoorn, J.C. (1999). Personality, self-efficacy and functioning in management teams: A contribution to validation. *International journal of selection and assessment*, 7, 56-49.
- Gellatly, I.R. & Irving, P.G. (2001). Personality, autonomy and contextual performance of managers. *Human performance*, 14, 231-245.
- Judge, T.A. & Bono, J.E. (2000). Five-factor model of personality and transformational leadership. *Journal of applied psychology*, 85, 751-765.

- Kickul, J. & Neuman, G. (2000). Emergent leadership behaviors: The function of personality and cognitive ability in determining teamwork performance and ksas. *Journal of business and psychology*, 15, 2-51.
- McCredie, H. & Shackleton, V. (2000). The unit general manager: a competency profile. *Personnel review*, 29, 106-114.
- Motowidlo, S.J., Burnett, J.R., Maczynski, J., Witkowski, S., Wojtachnio, A. & Chelapa, S. (1996). Predicting managerial job performance from personality ratings based on a structured interview: An international replication. *Polish Psychological Bulletin*,
- Orpen, C. (1996). Dependency as a moderator of the effects of networking behavior on managerial career success. *Journal of psychology*, 130, 245-247.
- Ployhart, R.E., Lim, B.C. & Chan, K.Y. (2001). Exploring relations between typical and maximum performance ratings and the five factor model of personality. *Personnel psychologie*, 54, 809-843.
- Robertson, I.T., Baron, H., Gibbons, P., MacIver, R. & Nyfield, G. (2000). Conscientiousness and managerial performance. *Journal of occupational and organizational psychology*, 73, 171-180.
- Stricker, L.J. & Rock, D. (1998). Assessing leadership potential with a biographical measure of personality traits. *International journal of selection and assessment*, 6, 164-184.
- Taggar, S., Hackett, R. & Saha, S. (1999). Leadership emergence in autonomous work teams: Antecedents and outcomes. *Personnel psychology*, 52, 899-926.
- Thomas, J.L., Dickson, M.W. & Bliese, P.D. (2001). Values predicting leader performance in the U.S. Army reserve officer training corps assessment Center: evidence for a personality-mediated model. *Leadership quarterly*, 12, 181-196.

Literaturgestützte Anforderungsanalyse

- Allan, P. (1981). Managers at work: A large scale study of the managerial job in New York City Government. *Academy of Management Journal*, 3, 613-619.
- Barnard, C.I. (1968). *The Function of the Executive*. Cambridge, Mass: Harvard University Press.

- Baugut, G. & Krüger, S. (1976). Unternehmensführung: Modelle - Strategien - Techniken. München: Opladen.
- Brooks, E. (1950). What Successful Executives Do. *Personnel*, 210-225.
- Fayol, H. (1929). Allgemeine und industrielle Verwaltung. Berlin.
- Goodman, R.D. (1968). A System Diagram of the Functions of a Manager. *California Management Review*, 4, 27-38.
- Haas, J.A., Porat, A.M. & Vaughan, J.A. (1969). Actual vs. ideal time allocations reported by managers: A study of managerial behavior. *Personnel Psychology*, 22, 61-75.
- Hinrichs, J.R. (1975). Where Has All the Time Gone? *Personnel*, 4, 44-49.
- Horne, J.H. & Lupton, N. (1975). The work activities of middle managers - an exploratory study. *Journal of Management Studies*, 2, 14-33.
- Koontz, H. & Weihrich, H. (1993). Management: A Global Perspektive. New Jersey: McGraw-Hill.
- Kotter, J.P. (1982). What effective managers really do. *Harvard Business Review*, 156-167.
- Lau, A.W., Newman, A.R. & Broedling, L.A. (1980). The nature of managerial work in the public sector. *Public Management Forum*, 513-520.
- Luthans, F. & Lockwood, D.L. (1984). Toward an observation system for measuring leader behavior in natural settings. In J.G. Hunt, D. Hosking, C.A. Schriesheim & R. Stewart (Hrsg.), *Leaders and managers: International perspectives on managerial behavior and leadership*. New York: Pergamon Press.
- Mahoney, T.A., Jerdee, T.H. & Carroll, S.J. (1965). The Job(s) of Management. *Industrial Relations*, 4(2), 97-110.
- Miner, J.B. (1973). *The Management Process – Theory, Research and Practice*. New York: Macmillan.
- Mintzberg, H. (1973). A New Look at the Chief Executive's Job. *Organizational Dynamics*, 21-30.
- Morse, J.J. & Wagner, F.R. (1978). Measuring the Process of Managerial Effectiveness. *Academy of Management Journal*, 21(1), 23-35.
- Newman (1963). *Administrative Action*. New Jersey: Englewood Cliffs.
- Parsons, T. (1960). *Structure and process in modern society*. New York.
- Stogdill, R.M. (1963). *Manual for the Leader Behavior Description Questionnaire - Form XII*. Columbus: Ohio State University.

- Terry, G.R. & Franklin, S.G. (1982). Principles of Management. Homewood.
- Urwick, L.F. (1956). The Patterns of Management. Minneapolis: University of Minnesota Press.
- Whitely, W. (1985). Managerial work behavior: An integration of results from two major approaches. *Academy of Management Journal*, 28 (2), 344-362.
- Wren, D.A. & Voich, D. (1984). Management: Process, Structure and Behavior. New York: Wiley.
- Yukl, G. (1989). Managerial leadership: A review of theory and research. *Journal of Management*, 15, 251-289.
- Yukl, G. (1994). Leadership in Organizations. New Jersey: Prentice Hall.
- Page (1978) Managerial Position Description Questionnaire.